Submitted to the Annals of Statistics

MULTIVARIATE DENSITY ESTIMATION UNDER SUP-NORM LOSS: ORACLE APPROACH, ADAPTATION AND INDEPENDENCE STRUCTURE

By Oleg Lepski

Université Aix–Marseille

The paper deals with the density estimation on $\mathbb{R}^d$ under sup-norm loss. We provide with fully data-driven estimation procedure and establish for it so called sup-norm oracle inequality. The proposed estimator allows to take into account not only approximation properties of the underlying density but eventual independence structure as well. Our results contain, as a particular case, the complete solution of the bandwidth selection problem in multivariate density model. Usefulness of the developed approach is illustrated by application to adaptive estimation over anisotropic Nikolskii classes.

1. Introduction. Let $(\Omega,\mathfrak{A},\mathrm{P})$ be a complete probability space and let $X_i=\big(X_{1,i},\dots X_{d,i}\big)$, $i\geq 1$, be the sequence of $\mathbb{R}^d$-valued *i.i.d.* random variables defined on $(\Omega,\mathfrak{A},\mathrm{P})$ and having the density f with respect to lebesgue measure. Furthermore, $\mathbb{P}^{(n)}_f$ denotes the probability law of $X^{(n)}=\big(X_1,\dots,X_n\big)$, $n\in\mathbb{N}^*$ and $\mathbb{E}^{(n)}_f$ is the mathematical expectation with respect to $\mathbb{P}_f$.

The objective is to estimate the density f and the quality of any estimation procedure, i.e. $X^{(n)}$-measurable mapping $\widehat{f}_n:\mathbb{R}^d\to\mathbb{L}_1(\mathbb{R}^d)$, is measured by *sup-norm risk* given by

$$R^{(q)}_n\big(\widehat{f},f\big)=\Big(\mathbb{E}^{(n)}_f\big\|\widehat{f}_n-f\big\|^q_\infty\Big)^{\frac{1}{q}},\;\; q\geq 1.$$

It is well-known that even asymptotically $(n\to\infty)$ the quality of estimation given by $R^{(q)}_n$ heavily depends on the dimension d. However, this asymptotics can be essentially improved if the underlying density possesses some special structure. Let us briefly discuss one of these possibilities which will be exploited in the sequel.

Introduce the following notations. Let $\mathcal{I}_d$ be the set of all subsets of $\{1,\dots,d\}$. For any $\mathbf{I}\in\mathcal{I}_d$ denote $x_{\mathbf{I}}=\{x_j\in\mathbb{R},\; j\in\mathbf{I}\}$, $\;\bar{\mathbf{I}}=\{1,\dots,d\}\setminus\mathbf{I}$ and let $|\mathbf{I}|=\mathrm{card}(\mathbf{I})$. Moreover for any function $g:\mathbb{R}^{|\mathbf{I}|}\to\mathbb{R}$ we denote $\|g\|_{\mathbf{I},\infty}=\sup_{x_{\mathbf{I}}\in\mathbb{R}^{|\mathbf{I}|}}|g(x_{\mathbf{I}})|$. Define also

$$f_{\mathbf{I}}(x_{\mathbf{I}})=\int_{\mathbb{R}^{|\bar{\mathbf{I}}|}}f(x)\mathrm{d}x_{\bar{\mathbf{I}}},\;\; x_{\mathbf{I}}\in\mathbb{R}^{|\mathbf{I}|}.$$

In accordance with this definition we put $f_{\mathbf{I}}\equiv 1,\;\; \mathbf{I}=\emptyset$. As we see $f_{\mathbf{I}}$ is the marginal density of $X_{\mathbf{I},1}:=\{X_{j,1},\; j\in\mathbf{I}\}$. Denote by $\mathfrak{P}$ the set of all partitions of $\{1,\dots,d\}$ completed by empty set $\emptyset$ and we will use $\bar{\emptyset}$ for $\{1,\dots,d\}$. For any density f let

$$\mathfrak{P}(f)=\Big\{\mathcal{P}\in\mathfrak{P}:\;\; f(x)=\prod_{\mathbf{I}\in\mathcal{P}}f_{\mathbf{I}}(x_{\mathbf{I}}),\;\;\forall x\in\mathbb{R}^d\Big\}.$$

AMS 2000 subject classifications: 62G07

Keywords and phrases: density estimation, oracle inequality, adaptation, upper function

First we note that $f \equiv f_{\bar{\emptyset}}$ and, therefore $\mathfrak{P}(f) \neq \emptyset$ since $\bar{\emptyset} \in \mathfrak{P}(f)$ for any f. Next, if $\mathcal{P} \in \mathfrak{P}(f)$ then $\{X_{\mathbf{I},1}, \mathbf{I} \in \mathcal{P}\}$ are independent random vectors. At last, if $X_{1,1}, \dots X_{d,1}$, are independent random variables then obviously $\mathfrak{P}(f) = \mathfrak{P}$.

Suppose now that there exists $\boldsymbol{\mathcal{P}} \neq \bar{\emptyset}$ such that $\boldsymbol{\mathcal{P}} \in \mathfrak{P}(f)$. If this partition is known we can proceed as follows. For any $\boldsymbol{I} \in \boldsymbol{\mathcal{P}}$ basing on observation $X_{\boldsymbol{I}}^{(n)}$ we estimate first the marginal density$f_{\mathbf{I}}$ by $\widehat{f}_{\mathbf{I},n}$ and then construct the estimator for joint density f as

$$\widehat{f}_n(x) = \prod_{\boldsymbol{I} \in \boldsymbol{\mathcal{P}}} \widehat{f}_{\boldsymbol{I},n}\left(x_{\boldsymbol{I}}\right).$$

One can expect (and we will see that our conjecture is true) that quality of estimation provided by this estimator will correspond not to the dimension d but to so-called effective dimension, which in our case is defined as $d(\boldsymbol{\mathcal{P}}) = \sup_{\boldsymbol{I} \in \boldsymbol{\mathcal{P}}} |\boldsymbol{I}|$. The main difficulty we meet trying to realize the latter construction is that the knowledge of $\boldsymbol{\mathcal{P}}$ is not available. Moreover, our structural hypothesis cannot be true in general, that is expressed formally by $\mathfrak{P}(f) = \{\bar{\emptyset}\}$. So, one of the problem we address in the present paper consists in adaptation to unknown configuration $\boldsymbol{\mathcal{P}} \in \mathfrak{P}(f)$.

We note however that even if $\boldsymbol{\mathcal{P}}$ is known, for instance, $\boldsymbol{\mathcal{P}} = \bar{\emptyset}$ the quality of an estimation procedure depends often on approximation properties of f or $\{\widehat{f}_{\boldsymbol{I},n},\ \boldsymbol{I} \in \boldsymbol{\mathcal{P}}\}$. So, our second goal is to construct an estimator which would mimic an estimator corresponding to the minimal, and therefore unknown, approximation error. Using modern statistical language our goal here is to mimic an oracle. It is important to emphasize that we would like to solve both aforementioned problem simultaneously. Let us now proceed with detailed consideration.

Collection of estimators. Let $\mathbf{K} : \mathbb{R} \to \mathbb{R}$ be a given function satisfying the following assumption.

Assumption 1. $\int \mathbf{K} = 1$, $\|\mathbf{K}\|_\infty < \infty$, *supp*$(\mathbf{K}) \subseteq [-1/2, 1/2]$, $\mathbf{K}$ *is symmetric, and*

$$\exists L > 0 : \quad |\mathbf{K}(t) - \mathbf{K}(s)| \leq L|t - s|, \quad \forall t, s \in \mathbb{R}.$$

Put for $\mathbf{I} \in \mathcal{I}_d$

$$K_{h_{\mathbf{I}}}(u) = V_{h_{\mathbf{I}}}^{-1} \prod_{j \in \mathbf{I}} \mathbf{K}\big(u_j/h_j\big), \qquad V_{h_{\mathbf{I}}} = \prod_{j \in \mathbf{I}} h_j.$$

For two vectors u, v here and later u/v denotes coordinate-vise division. We will use the notation $V_h = \prod_{j=1}^d h_j$ instead of $V_{h_{\mathbf{I}}}$ then $\mathbf{I} = \{1, \dots, d\}$. Denote also $\mathrm{k}_m = \|\mathbf{K}\|_m$, $m = \{1, \infty\}$.

For any $p \geq 1$ let $\gamma_p : \mathbb{N}^* \times \mathbb{R}_+ \to \mathbb{R}_+$ be the function whose explicit expression is given in Section 2.3 (it has quite cumbersome expression and it is not convenient for us to present it right now).

Introduce the notations (remind that q is the quantity involved in the definition of the risk)

$$\mathcal{H}_n = \big\{h \in (0,1]^d : \ nV_h \geq (\mathfrak{a}^*)^{-1} \ln(n)\big\}, \qquad \mathfrak{a}^* = \inf_{\mathbf{I} \in \mathcal{I}_d} \big[2\gamma_{2q}\big(|\mathbf{I}|, \mathrm{k}_\infty\big)\big]^{-2}$$

and for any $\mathbf{I} \in \mathcal{I}_d$ and $h \in \mathcal{H}_n$ consider kernel estimator

$$\widetilde{f}_{h_{\mathbf{I}}}(x_{\mathbf{I}}) = n^{-1} \sum_{i=1}^n K_{h_{\mathbf{I}}}\left(X_{\mathbf{I},i} - x_{\mathbf{I}}\right).$$

Introduce the family of estimators

$$\mathfrak{F}(\mathfrak{P}) = \left\{\widehat{f}_{h,\mathcal{P}}(x) = \prod_{\mathbf{I} \in \mathcal{P}} \widetilde{f}_{h_{\mathbf{I}}}\big(x_{\mathbf{I}}\big), \ \ x \in \mathbb{R}^d, \ \ \mathcal{P} \in \mathfrak{P}, \ h \in \mathcal{H}_n\right\}.$$

In particular, $\widehat{f}_{h,\bar{\emptyset}}(x) = n^{-1}\sum_{i=1}^{n} K_h\left(X_i - x\right),\ x \in \mathbb{R}^d$, is the Parzen-Rosenblatt estimator (Parzen (1962), Rosenblatt (1956)) with kernel K and multi-bandwidth h. Our goal is to propose a data-driven selection from the family $\mathfrak{F}(\mathfrak{P})$.

The estimation of a probability density is the subject of the vast literature. We do not pretend here to provide with complete overview and only present the results relevant in context of the considered problems. Minimax and minimax adaptive density estimation with $\mathbb{L}_s$–risks was considered in Bretagnolle and Huber (1979), Ibragimov and Khasminskii (1980, 1981), Devroye and Györfi (1985), Efroimovich (1986, 2008), Hasminskii and Ibragimov (1990), Donoho et al. (1996), Golubev (1992), Kerkyacharian, Picard and Tribouley (1996), Juditsky and Lambert–Lacroix (2004), Rigollet (2006), Mason (2009), Reynaud-Bouret, Rivoirard and Tuleau-Malot (2011) and Akakpo (2012), where further references can be found. Oracle inequalities for $\mathbb{L}_s$–risks for $s = 1$ and $s = 2$ were established in Devroye and Lugosi (1996, 1997, 2001), Massart (2007)[Chapter 7], Samarov and Tsybakov (2007), Rigollet and Tsybakov (2007) and Birgé (2008). The last cited paper contains a detailed discussion of recent developments in this area. Bandwidth selection problem in the density estimation on $\mathbb{R}^d$ with $\mathbb{L}_s$–risks for any $1 \le s < \infty$ was studied in Goldenshluger and Lepski (2011). The oracle inequalities obtained there were used for deriving adaptive minimax results over the collection of anisotropic Nikolskii classes.

The adaptive estimation under sup-norm loss was initiated in Lepski (1991, 1992) and continued in Tsybakov (1998) in the framework of gaussian white noise model. Then, it was developed for anisotropic functional classes in Bertin (2005). The adaptive estimation of a probability density on $\mathbb{R}$ in sup-norm was the subject of recent papers Giné and Nickl (2009, 2010).

Organization of the paper. In Section 2 we present data-driven selection procedure from $\mathfrak{F}(\mathfrak{P})$ and establish for it sup-norm oracle inequality. Section 3 is devoted to the adaptive estimation over the collection of anisotropic Nikolskii classes of functions. The proof of main results are given in Section 4 and technical lemmas are proven in Appendix.

2. Oracle inequality.

Let $\mathcal{P} \in \mathfrak{P}$ be fixed and define for any $h, \eta \in \mathcal{H}_n$ and any $\mathbf{I} \in \mathcal{P}$

$$\widetilde{f}_{h_{\mathbf{I}},\eta_{\mathbf{I}}}(x_{\mathbf{I}}) = n^{-1}\sum_{i=1}^{n}\left[K_{h_{\mathbf{I}}} \star K_{\eta_{\mathbf{I}}}\right]\left(X_{\mathbf{I},i} - x_{\mathbf{I}}\right),$$

where $\left[K_{h_{\mathbf{I}}} \star K_{\eta_{\mathbf{I}}}\right] = \prod_{j\in\mathbf{I}}\left[\mathbf{K}_{h_j} * \mathbf{K}_{\eta_j}\right]$ and $\left[\mathbf{K}_{h_j} * \mathbf{K}_{\eta_j}\right](z) = \int_{\mathbb{R}} \mathbf{K}_{h_j}(u-z)\mathbf{K}_{\eta_j}(u)\mathrm{d}u,\ \ z \in \mathbb{R}$. As we see " $\star$ " is the convolution operator on $\mathbb{R}^{|\mathbf{I}|}$. Define

$$\mathbf{f}_n = \sup_{h\in\mathcal{H}_n}\sup_{\mathbf{I}\in\mathcal{I}_d}\left\|n^{-1}\sum_{i=1}^{n}\left|K_{h_{\mathbf{I}}}\left(X_{\mathbf{I},i} - \cdot\right)\right|\right\|_{\mathbf{I},\infty}, \quad \bar{\mathbf{f}}_n = 1 \vee 2\mathbf{f}_n$$

$$\widehat{A}_n(h,\mathcal{P}) = \sqrt{\frac{\bar{\mathbf{f}}_n \ln(n)}{nV(h,\mathcal{P})}}, \quad V(h,\mathcal{P}) = \inf_{\mathbf{I}\in\mathcal{P}} V_{h_{\mathbf{I}}}.$$

Let us endow the set $\mathfrak{P}$ with the operation "$\diamond$" putting for any $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$\mathcal{P} \diamond \mathcal{P}' = \left\{\mathbf{I} \cap \mathbf{I}' \ne \emptyset,\ \ \mathbf{I} \in \mathcal{P},\ \mathbf{I}' \in \mathcal{P}'\right\} \in \mathfrak{P}.$$

Introduce for any $h, \eta \in \mathcal{H}_n$ and any $\mathcal{P}, \mathcal{P}'$ the estimator

$$\widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x) = \prod_{\mathbf{I}^\diamond\in\mathcal{P}\diamond\mathcal{P}'}\widetilde{f}_{h_{\mathbf{I}^\diamond},\eta_{\mathbf{I}^\diamond}}(x_{\mathbf{I}^\diamond}),\quad x \in \mathbb{R}^d.$$

Set finally $\Lambda = \sup_{\mathcal{P}\in\mathfrak{P}}\sup_{\mathbf{I}\in\mathcal{P}}\gamma_{2q}\left(|\mathbf{I}|, \mathrm{k}_\infty\right)$ and let $\lambda = \Lambda d\left(\bar{\mathbf{f}}_n\right)^{\lfloor d^2/4\rfloor+1}$.

2.1. *Selection procedure.* For any $\mathcal{P} \in \mathfrak{P}$ and $h \in \mathcal{H}_n$ set

$$\widehat{\Delta}_n(h, \mathcal{P}) = \sup_{\eta \in \mathcal{H}_n} \sup_{\mathcal{P}' \in \mathfrak{P}} \left[\left\| \widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widehat{f}_{\eta,\mathcal{P}'} \right\|_\infty - \lambda \widehat{A}_n\big(\eta, \mathcal{P}'\big) \right]_+,$$

and let $\widehat{h}$ and $\widehat{\mathcal{P}}$ be defined as follows.

$$\widehat{\Delta}_n\big(\widehat{h}, \widehat{\mathcal{P}}\big) + \lambda \widehat{A}_n\big(\widehat{h}, \widehat{\mathcal{P}}\big) = \inf_{h \in \mathcal{H}_n} \inf_{\mathcal{P} \in \mathfrak{P}} \left[\widehat{\Delta}_n\big(h, \mathcal{P}\big) + \lambda \widehat{A}_n\big(h, \mathcal{P}\big) \right].$$

Our final estimator is $\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}(x), \;\; x \in \mathbb{R}^d$.

Existence and measurability. Let us briefly discuss the existence of the proposed estimator as well as its the measurability with respect to the σ-algebra generated by $X^{(n)}$. First, we note that all considered in the paper random fields have continuous trajectories on $\mathcal{H}_n \times \mathbb{R}^d$ in the topology generated by supremum norm. It is guaranteed by Assumption 1. Since $\mathcal{H}_n$ is totally bounded and $\mathbb{R}^d$ can be covered by a countable collection of totally bounded sets, any supremum over $\mathcal{H}_n \times \mathbb{R}^d$ of considered random fields will be $X^{(n)}$-measurable. In particular, $\bar{\mathbf{f}}_n$ and

$$\widehat{\Delta}_n(h, \mathcal{P}, \mathcal{P}') := \sup_{\eta \in \mathcal{H}_n} \sup_{\mathcal{P}' \in \mathfrak{P}} \left[\left\| \widehat{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widehat{f}_{\eta,\mathcal{P}'} \right\|_\infty - \lambda \widehat{A}_n\big(\eta, \mathcal{P}'\big) \right]_+, \quad \mathcal{P}, \mathcal{P}' \in \mathfrak{P}, \; h \in \mathcal{H}_n.$$

Since, $\mathfrak{P}$ is finite, we conclude that $\widehat{\Delta}_n(h, \mathcal{P})$ is $X^{(n)}$-measurable for any $\mathcal{P} \in \mathfrak{P}$ and any $h \in \mathcal{H}_n$. Assumption 1 implies also that $\widehat{\Delta}_n(\cdot, \mathcal{P})$ and $\widehat{A}_n\big(\cdot, \mathcal{P}\big)$ are continuous on $\mathcal{H}_n$ for any $\mathcal{P}$. Since $\mathcal{H}_n$ is a compact subset of $\mathbb{R}^d$ we conclude that $\widehat{h}(\mathcal{P}) \in \mathcal{H}_n$ and $X^{(n)}$-measurable for any $\mathcal{P} \in \mathfrak{P}$, Jennrich (1969), where $\widehat{h}(\mathcal{P}) = \inf_{h \in \mathcal{H}_n} \left[\widehat{\Delta}_n\big(h, \mathcal{P}\big) + \lambda \widehat{A}_n\big(h, \mathcal{P}\big) \right]$. Since $\mathfrak{P}$ is finite we conclude that $(\widehat{h}, \widehat{\mathcal{P}}) \in \mathcal{H}_n \times \mathfrak{P}$ is $X^{(n)}$-measurable.

2.2. *Main result.* Let $\mathbf{f} > 0$ be a given number and introduce the following set of densities

$$\mathbf{F}(\mathbf{f}) = \left\{ f : \; \sup_{\mathbf{I} \in \mathcal{I}_d} \|f_{\mathbf{I}}\|_\infty \leq \mathbf{f} \right\}.$$

With any density $f \in \mathbf{F}(\mathbf{f})$, any $h \in (0,1]^d$ and $\mathbf{I} \in \mathcal{I}_d$ associate the quantity

$$b_{h_{\mathbf{I}}} := \left\| \int_{\mathbb{R}^{|\mathbf{I}|}} K_{h_{\mathbf{I}}}\big(t_{\mathbf{I}} - \cdot\big) \big[f_{\mathbf{I}}\big(t_{\mathbf{I}}\big) - f_{\mathbf{I}}(\cdot) \big] \mathrm{d}t_{\mathbf{I}} \right\|_{\mathbf{I},\infty},$$

which can be view as the approximation error of $f_{\mathbf{I}}$.

For any $h \in \mathcal{H}_n$ and $\mathcal{P} \in \mathfrak{P}$ set $B\big(h, \mathcal{P}\big) = \sup_{\mathcal{P}'} \sup_{\mathbf{I} \in \mathcal{P} \diamond \mathcal{P}'} \|b_{h_{\mathbf{I}}}\|_{\mathbf{I},\infty}$ and introduce the quantity

$$\mathfrak{R}_n(f) = \inf_{h \in \mathcal{H}_n} \inf_{\mathcal{P} \in \mathfrak{P}(f)} \left(B\big(h, \mathcal{P}\big) + \sqrt{\frac{\ln(n)}{nV(h, \mathcal{P})}} \right).$$

Theorem 1. *Let Assumption 1 be fulfilled. Then for any $q \geq 1$ and any $0 < \mathbf{f} < \infty$ there exist $\mathbf{C_1}\big(q, d, \mathbf{K}, \mathbf{f}\big)$ and $\mathbf{C_2}\big(q, d, \mathbf{K}, \mathbf{f}\big)$ such that for any $f \in \mathbf{F}(\mathbf{f})$ and any $n \geq 3$*

$$\left(\mathbb{E}_f \big\| \widehat{f}_{\widehat{h},\widehat{\mathcal{P}}} - f \big\|_\infty^q \right)^{\frac{1}{q}} \leq \mathbf{C_1}\big(q, d, \mathbf{K}, \mathbf{f}\big) \mathfrak{R}_n(f) + \mathbf{C_2}\big(q, d, \mathbf{K}, \mathbf{f}\big) n^{-1/2}.$$

The explicit expression of $\mathbf{C_1}\big(q, d, \mathbf{K}, \mathbf{f}\big)$ and $\mathbf{C_2}\big(q, d, \mathbf{K}, \mathbf{f}\big)$ can be found in the proof of the theorem.

Discussion. Let us briefly discuss the assertion of Theorem 1. We start with the following simple observation. Let $\bar{\mathfrak{P}}$ be an arbitrary subset of $\mathfrak{P}$ containing $\bar{\emptyset}$. If our selection rule run $\bar{\mathfrak{P}}$ instead of $\mathfrak{P}$ then the result of the theorem remains valid if one replaces the quantity $\mathfrak{R}_n(f)$ by

$$\bar{\mathfrak{R}}(f) = \inf_{h\in\mathcal{H}_n} \inf_{\mathcal{P}\in\bar{\mathfrak{P}}(f)} \left(B\big(h,\mathcal{P}\big) + \sqrt{\frac{\ln(n)}{nV(h,\mathcal{P})}} \right),$$

where $\bar{\mathfrak{P}}(f) = \mathfrak{P}(f)\cap\bar{\mathfrak{P}}$. The reason of considering $\bar{\mathfrak{P}}$ instead of $\mathfrak{P}$ is explained by the fact that the cardinality of $\mathfrak{P}$ (Bell number) grows as $(d/\ln(d))^d$. Therefore, for large dimension our procedure is not practically feasible in view of huge amount of comparisons to be done. On the other hand if d is large the consideration of all partitions is not reasonable. Indeed, even theoretically the best attainable trade-off between approximation and stochastic errors corresponds to the effective dimension defined as $d^*(f) = \inf_{\mathcal{P}\in\mathfrak{P}(f)} \sup_{\mathbf{I}\in\mathcal{P}} |\mathbf{I}|$. Of course $d^*(f) \leq d$ but if it is proportional for example to d then we will not win much for reasonable sample size. The suitable strategy in the case of large dimension consists in considering only partitions satisfying $\sup_{\mathbf{I}\in\mathcal{P}} |\mathbf{I}| \leq d_0$, where d_0 is chosen in accordance with d and the number of observation. In particular one can consider $\bar{\mathfrak{P}}$ containing only 2 elements namely $\bar{\emptyset}$ and $\big(\{1\},\{2\},\dots\{d\}\big)$. It corresponds to the hypotheses that we observe vectors with independent components.

Of course the consideration of $\bar{\mathfrak{P}}$ instead of $\mathfrak{P}$ has a price to pay. It is possible that $\mathfrak{P}(f)\cap\bar{\mathfrak{P}} = \bar{\emptyset}$ although $\mathfrak{P}(f)$ contains the elements besides $\bar{\emptyset}$. However even in this case, where structural hypothesis fails or is not taken into account ($\bar{\mathfrak{P}} = \{\bar{\emptyset}\}$), our estimator solves completely the bandwidths selection problem in multivariate density model under sup-norm loss.

We finish this discussion with the following remark concerning the proof of Theorem 1.

Remark 1. *Our selection rule is based on computation of upper functions for some special type of random processes and the main ingredient of the proof of Theorem 1 is exponential inequality related to them. Corresponding results may have an independent interest and Section 4.1 is devoted to this topic. In particular the function γ_p involved in the construction of our selection rule and which we present below comes from this consideration.*

2.3. *Quantity γ_p.* For any $a>0$, $p\geq 1$ and $s\in\mathbb{N}^*$ introduce

$$\begin{aligned}
\gamma_p(s,a) &= 4e\sqrt{2s\tau_p(s,a)\left[a+(3L/2)(a)^{s-1}\right]} + (16e/3)\left(s\left[a+(3L/2)a^{s-1}\right]\vee 8a\right)\tau_p(s,a);\\
\tau_p(s,a) &= s\big(234s\delta_*^{-2}+6.5p+5.5\big)\ln(2) + s(2p+3) + \left[108s\delta_*^{-2}\big|\log(a)\big| + 36C_s+1\right][\ln(3)]^{-1}.
\end{aligned}$$

Here δ_* is the smallest solution of the equation $8\pi^2\delta\big(1+[\ln\delta]^2\big)=1$, $C_s = C_s^{(1)}+C_s^{(2)}$ and

$$C_s^{(1)} = s\sup_{\delta>\delta_*}\delta^{-2}\left\{\left[1+\ln\left(\frac{9216(s+1)\delta^2}{[\phi(\delta)]^2}\right)\right]_+ + 1.5\left[\log_2\left\{\left(\frac{4608(s+1)\delta^2}{[\phi(\delta)]^2}\right)\right\}\right]_+\right\};$$

$$C_s^{(2)} = s\sup_{\delta>\delta_*}\delta^{-1}\left\{\left[1+\ln\left(\frac{9216(s+1)\delta}{\phi(\delta)}\right)\right]_+ + 1.5\left[\log_2\left\{\left(\frac{4608(s+1)\delta}{\phi(\delta)}\right)\right\}\right]_+\right\},$$

where $\phi(\delta) = (6/\pi^2)\big(1+[\ln\delta]^2\big)^{-1},\ \delta>0$.

3. Adaptive Estimation. In this section we illustrate the use of the oracle inequality proved in Theorem 1 for the derivation of adaptive rate optimal density estimators.

We start with the definition of the *anisotropic Nikol'skii class of functions* on $\mathbb{R}^s$, $s \geq 1$, and later on $\mathbf{e}_1, \dots \mathbf{e}_s$, denotes the canonical basis in $\mathbb{R}^s$.

DEFINITION 1. *Let $r = (r_1, \dots, r_s), r_i \in [1, \infty]$, $\alpha = (\alpha_1, \dots, \alpha_s)$, $\alpha_i > 0$, and $Q = (Q_1, \dots, Q_s)$, $Q_i > 0$. A function $g : \mathbb{R}^s \to \mathbb{R}$ belongs to the anisotropic Nikol'ski class $\mathbb{N}_{r,s}(\alpha, Q)$ of functions if*

$$\|D_i^k g\|_{r_i} \leq Q_i, \quad \forall k = \overline{0, \lfloor \alpha_i \rfloor}, \ \ \forall i = \overline{1, s};$$

$$\Big\| D_i^{\lfloor \alpha_i \rfloor} g\big(\,\cdot + t\mathbf{e}_i\big) - D_i^{\lfloor \alpha_i \rfloor} g\big(\,\cdot\,\big) \Big\|_{r_i} \leq Q_i |t|^{\alpha_i - \lfloor \alpha_i \rfloor}, \quad \forall t \in \mathbb{R}, \ \ \forall i = \overline{1, s}.$$

Here $D_i^k f$ denotes the kth order partial derivative of f with respect to the variable t_i, and $\lfloor \alpha_i \rfloor$ is the largest integer strictly less than α_i.

The functional classes $\mathbb{N}_{r,s}(\alpha, Q)$ were considered in approximation theory by Nikol'skii; see, e.g., Nikol'skii (1977). Minimax estimation of densities from the class $\mathbb{N}_{r,s}(\alpha, Q)$ was considered in Ibragimov and Khasminskii (1981). We refer also to Kerkyacharian, Lepski and Picard (2001, 2007), where the problem of adaptive estimation over a scale of classes $\mathbb{N}_{r,s}(\alpha, Q)$ was treated for the Gaussian white noise model.

Our goal now is to introduce the scale of functional classes of d-variate probability densities taking into account the independence structure. It implies in particular that we will need to estimate not only the density itself but all marginal densities as well. It is easily seen that if $f \in \mathbb{N}_{p,d}(\beta, \mathcal{L})$ and additionally f is compactly supported then $f_{\mathbf{I}} \in \mathbb{N}_{p_{\mathbf{I}}, |\mathbf{I}|}\big(\beta_{\mathbf{I}}, \overline{\mathcal{L}}_{\mathbf{I}}\big)$ for any $\mathbf{I} \in \mathcal{I}_d$, where $\overline{\mathcal{L}} = \mathrm{c}\mathcal{L}$ and $\mathrm{c} > 0$ is a numerical constant. However if $\mathrm{supp}(f) = \mathbb{R}^d$ the latter assertion is not true in general. The assumption $f \in \mathbb{N}_{p,d}(\beta, \mathcal{L})$ does not even guarantee that $f_{\mathbf{I}}$ is bounded on $\mathbb{R}^{|\mathbf{I}|}$. It explains the introduction of the following anisotropic classes of densities.

Let $p = (p_1, \dots, p_d), p_i \in [1, \infty]$, $\beta = (\beta_1, \dots, \beta_d)$, $\beta_i > 0$, $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_d)$, $\mathcal{L}_i > 0$.

DEFINITION 2. *A probability density $f : \mathbb{R}^d \to \mathbb{R}_+$ belongs to the class $\overline{\mathbb{N}}_{p,d}\big(\beta, \mathcal{L}\big)$ if*

$$f_{\mathbf{I}} \in \mathbb{N}_{p_{\mathbf{I}}, |\mathbf{I}|}\big(\beta_{\mathbf{I}}, \mathcal{L}_{\mathbf{I}}\big), \quad \forall \mathbf{I} \in \mathcal{I}_d.$$

Introduce finally the collection of functional classes taking into account the smoothness of the underlying density and the independence structure simultaneously.

Let $\big(\beta, p, \mathcal{P}\big) \in (0, \infty)^d \times [1, \infty]^d \times \mathfrak{P}$ and $\mathcal{L} \in (0, \infty)^d$ be fixed. Introduce

$$\mathbf{N}_{p,d}\big(\beta, \mathcal{L}, \mathcal{P}\big) = \bigg\{ f(x) \in \overline{\mathbb{N}}_{p,d}\big(\beta, \mathcal{L}\big) : \ \ f(x) = \prod_{\mathbf{I} \in \mathcal{P}} f_{\mathbf{I}}(x_{\mathbf{I}}), \ \ \forall x \in \mathbb{R}^d \bigg\}.$$

For any $\big(\beta, p, \mathcal{P}\big) \in (0, \infty)^d \times [1, \infty]^d \times \mathfrak{P}$ define

$$\Upsilon\big(\beta, p, \mathcal{P}\big) = \inf_{\mathbf{I} \in \mathcal{P}} \gamma_{\mathbf{I}}(\beta, p), \qquad \gamma_{\mathbf{I}}\big(\beta, p\big) = \frac{1 - \sum_{j \in \mathbf{I}} \frac{1}{\beta_j p_j}}{\sum_{j \in \mathbf{I}} \frac{1}{\beta_j}}.$$

We will see that the quantity $\Upsilon\big(\beta, p, \mathcal{P}\big)$ can be view as "effective smoothness index" related to independence structure hypothesis and to the estimation under sup-norm loss.

Theorem 2. *For any $\big(\beta,p,\mathcal{P}\big)\in(0,\infty)^d\times[1,\infty]^d\times\mathfrak{P}$ such that $\Upsilon\big(\beta,p,\mathcal{P}\big)>0$ and any $\mathcal{L}\in(0,\infty)^d$*

$$\liminf_{n\to\infty}\inf_{\widehat{f}_n}\sup_{f\in\mathbf{N}_{p,d}\left(\beta,\mathcal{L},\mathcal{P}\right)}\left(\mathbb{E}^{(n)}_f\left[\varphi_n^{-1}(\beta,p,\mathcal{P})\big\|\widehat{f}_n-f\big\|_\infty\right]^q\right)^{\frac{1}{q}}>0,\qquad\varphi_n(\beta,p,\mathcal{P})=\left(\frac{\ln n}{n}\right)^{\frac{\Upsilon}{2\Upsilon+1}}.$$

where $\Upsilon=\Upsilon\big(\beta,p,\mathcal{P}\big)$ and infimum is taken over all possible estimators.

Our goal is to prove that the estimation quality provided by $\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}$ on $\mathbf{N}_{p,d}\big(\beta,\mathcal{L},\mathcal{P}\big)$ coincides up to numerical constant with optimal decay of minimax risk $\varphi_n\big(\beta,p,\mathcal{P}\big)$ whenever the value of nuisance parameter $\big\{\beta,p,\mathcal{P},\mathcal{L}\big\}$. It means that this estimator is optimally adaptive over the scale of considered functional classes. We would like to emphasize that not only the couple $(\beta,\mathcal{L})$ is unknown that is typical in frameworks of adaptive estimation but also the index p of norms where the smoothness is measured. At last, our estimator adapts automatically to unknown independence structure.

Theorem 3. *Let $\mathbf{K}$ satisfy Assumption 1 and suppose additionally that for some $\mathfrak{b}>2$*

$$\int_{\mathbb{R}}u^m\mathbf{K}(u)\mathrm{d}u=0,\qquad\forall m=\overline{2,\mathfrak{b}}.\tag{3.1}$$

Then for any $\big(\beta,p,\mathcal{P}\big)\in(0,\mathfrak{b}]^d\times[1,\infty]^d\times\mathfrak{P}$ such that $\Upsilon\big(\beta,p,\mathcal{P}\big)>0$ and any $\mathcal{L}\in(0,\infty)^d$

$$\limsup_{n\to\infty}\sup_{f\in\mathbf{N}_{p,d}\left(\beta,\mathcal{L},\mathcal{P}\right)}\left(\mathbb{E}^{(n)}_f\left[\varphi_n^{-1}(\beta,p,\mathcal{P})\big\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}-f\big\|_\infty\right]^q\right)^{\frac{1}{q}}<\infty.$$

We want to emphasize that the extra-parameter $\mathfrak{b}$ can be arbitrary but *a priory* chosen. Note that the condition (3.1) of the theorem is fulfilled with $m=1$ as well since $\mathbf{K}$ is symmetric.

We remark also that for any given $\big(\beta,p,\mathcal{P}\big)\in(0,\mathfrak{b}]^d\times[1,\infty]^d\times\mathfrak{P}$, satisfying $\Upsilon\big(\beta,p,\mathcal{P}\big)>0$, one can find $\mathbf{f}=\mathbf{f}\big(\beta,p,\mathcal{P}\big)$ such that $f\in\mathbf{N}_{p,d}\big(\beta,\mathcal{L},\mathcal{P}\big)$ implies that $f\in\mathbf{F}(\mathbf{f})$. It makes possible the application of Theorem 1.

4. Proofs. We start this section with the computation of upper functions for kernel estimation process being one of main tools in the proof of Theorem 1.

4.1. *Upper functions for kernel estimation process .* Let $s\in\mathbb{N}^*$ and let $Y_j, j\geq1$, be $\mathbb{R}^s$-valued i.i.d. random vectors defined on a complete probability space $(\Omega,\mathfrak{A},\mathrm{P})$ and having the density $\mathbf{g}$ with respect to the Lebesgue measure. Later on $\mathbb{P}^{(n)}_{\mathbf{g}}$ denotes the law of $Y_1,\dots,Y_n, n\in\mathbb{N}^*$, and $\mathbb{E}^{(n)}_{\mathbf{g}}$ is mathematical expectation with respect to $\mathbb{P}^{(n)}_{\mathbf{g}}$.

Let $\mathbf{M}:\mathbb{R}\to\mathbb{R}$ be a given symmetric function and for any $r\in(0,1]^s$ set as previously

$$M_r(\cdot)=\prod_{l=1}^s r_l^{-1}\mathbf{M}(\cdot/r_l),\qquad V_r=\prod_{l=1}^s r_l.$$

Denote also $\mathrm{m}_m=\|\mathbf{M}\|_m,\ m=\{1,\infty\}$. For any $y\in\mathbb{R}^s$ consider the family of random fields

$$\chi_r(y)=n^{-1}\sum_{j=1}^n\Big\{M_r\left(Y_j-y\right)-\mathbb{E}^{(n)}_{\mathbf{g}}\Big[M_r\left(Y_j-y\right)\Big]\Big\},\ \ r\in\widetilde{\mathcal{R}}_n(s):=\left\{r\in(0,1]^s:\ \ nV_r\geq\ln(n)\right\}.$$

For any $r\in(0,1]^s$ set $G(r)=\sup_{y\in\mathbb{R}^s}\int_{\mathbb{R}^s}|M_r(x-y)|\mathbf{g}(x)\mathrm{d}x$ and let $\bar{G}(r)=1\vee G(r)$.

Proposition 1. *Let* $\mathbf{M}$ *satisfy Assumption 1. Then for any* $n\geq 3$ *and any* $p\geq 1$

$$\mathbb{E}^{(n)}_{\mathbf{g}}\left\{\sup_{r\in\widetilde{\mathcal{R}}_n(s)}\left[\big\|\chi_r\big\|_\infty-\gamma_p\big(s,\mathrm{m}_\infty\big)\sqrt{\frac{\bar{G}(r)\ln(n)}{nV_r}}\right]\right\}^p_+\leq c_1(p,s)\big[1\vee \mathrm{m}_1^s\|\mathbf{g}\|_\infty\big]^{\frac{p}{2}}n^{-\frac{p}{2}}+c_2(p,s)n^{-p},$$

where $c_1(p,s)=2^{7p/2+5}3^{p+5s+4}\Gamma(p+1)\pi^p\big(s,\mathrm{m}_\infty\big)$ *and* $c_2(p,s)=2^{p+1}3^{5s}$.

The function $\pi:\mathbb{N}^*\times\mathbb{R}_+:\to\mathbb{R}$ is given by

$$\pi(s,a)=\big(\sqrt{a}\vee a\big)\left(\sqrt{2es\left[1+(3L/2)a^{s-2}\right]}\vee\left[(2e/3)\left(s\left[1+(3L/2)a^{s-2}\right]\vee 8\right)\right]\right).$$

In view of trivial inequality

$$\big\|\chi_r\big\|_\infty\leq\gamma_p\big(s,\mathrm{m}_\infty\big)\sqrt{\frac{\bar{G}(r)\ln(n)}{nV_r}}+\left(\big\|\chi_r\big\|_\infty-\gamma_p\big(s,\mathrm{m}_\infty\big)\sqrt{\frac{\bar{G}(r)\ln(n)}{nV_r}}\right)_+$$

we come to the following corollary of Proposition 1.

Corollary 1. *Let* $\mathbf{M}$ *satisfy Assumption 1. Then for any* $n\geq 3$ *and any* $p\geq 1$

$$\left(\mathbb{E}^{(n)}_{\mathbf{g}}\left\{\sup_{r\in\widetilde{\mathcal{R}}_n(s)}\big\|\chi_r\big\|_\infty\right\}^p\right)^{\frac{1}{p}}\leq\big[1\vee \mathrm{m}_1^s\|\mathbf{g}\|_\infty\big]^{\frac{1}{2}}\left[\gamma_p\big(s,\mathrm{m}_\infty\big)+\big\{c_1(p,s)+c_2(p,s)\big\}^{\frac{1}{p}}n^{-1/2}\right].$$

Consider now the following family of random processes: for any $y\in\mathbb{R}^s$

$$\Upsilon_r(y)=n^{-1}\sum_{j=1}^n\big|M_r\left(Y_j-y\right)\big|,\quad r\in\widetilde{\mathcal{R}}^{(\mathfrak{a})}_n(s):=\left\{r\in(0,1]^s:\ \ nV_r\geq\mathfrak{a}^{-1}\ln(n)\right\},$$

where we have put $\mathfrak{a}=\big[2\gamma_p\big(s,\mathrm{m}_\infty\big)\big]^{-2}$.

Proposition 2. *Let* $\mathbf{M}$ *satisfy Assumption 1. Then for any* $n\geq 3$ *and any* $p\geq 1$

$$\mathbb{E}^{(n)}_{\mathbf{g}}\left\{\sup_{r\in\mathcal{R}^{(\mathfrak{a})}_n(s)}\left[1\vee\|\Upsilon_r\|_\infty-(3/2)\bar{G}(r)\right]\right\}^p_+\leq c_1(p,s)\big[1\vee \mathrm{m}_1^s\|\mathbf{g}\|_\infty\big]^{\frac{p}{2}}n^{-\frac{p}{2}}+c_2(p,s)n^{-p};$$

$$\mathbb{E}^{(n)}_{\mathbf{g}}\left\{\sup_{r\in\mathcal{R}^{(\mathfrak{a})}_n(s)}\left[\bar{G}(r)-2\left(1\vee\|\Upsilon_r(\cdot)\|_\infty\right)\right]\right\}^p_+\leq c_1'(p,s)\big[1\vee \mathrm{m}_1^s\|\mathbf{g}\|_\infty\big]^{\frac{p}{2}}n^{-\frac{p}{2}}+c_2'(p,s)n^{-p},$$

where $c_1'(p,s)=2^pc_1(p,s)$ *and* $c_2'(p,s)=2^{2p+1}3^{5s}$.

4.2. *Proof of Theorem 1.* We start the proof of the theorem with auxiliary results used in the sequel. Whose proofs are given in Appendix.

4.2.1. *Auxiliary results.* Introduce the following notations. For any $\mathbf{I}\in\mathcal{I}_d$ set

$$s_{h_{\mathbf{I}}}(\cdot)=\int_{\mathbb{R}^{|\mathbf{I}|}}K_{h_{\mathbf{I}}}\big(t_{\mathbf{I}}-\cdot\big)f_{\mathbf{I}}\big(t_{\mathbf{I}}\big)\mathrm{d}t_{\mathbf{I}},\qquad s^*_{h_{\mathbf{I}},\eta_{\mathbf{I}}}(\cdot)=\int_{\mathbb{R}^{|\mathbf{I}|}}\left[K_{h_{\mathbf{I}}}\star K_{\eta_{\mathbf{I}}}\right]\big(t_{\mathbf{I}}-\cdot\big)f_{\mathbf{I}}\big(t_{\mathbf{I}}\big)\mathrm{d}t_{\mathbf{I}};$$

Lemma 1. *For any $\mathbf{I}\in\mathcal{I}_d$ and any $h,\eta\in(0,1]^{|\mathbf{I}|}$ one has*

$$\big\|s^*_{h_{\mathbf{I}},\eta_{\mathbf{I}}}-s_{\eta_{\mathbf{I}}}\big\|_{\mathbf{I},\infty}\le \mathrm{k}_1^d b_{h_{\mathbf{I}}}.$$

For any $h\in(0,1]^d$ and any $\mathcal{P}\in\mathfrak{P}$ set

$$A_n(h,\mathcal{P})=\sqrt{\frac{\bar{s}_n\ln(n)}{nV(h,\mathcal{P})}},\quad \bar{s}_n=1\vee\sup_{h\in\mathcal{H}_n}\sup_{\mathbf{I}\in\mathcal{I}_d}\left\|\int_{\mathbb{R}^{|\mathbf{I}|}}\big|K_{h_{\mathbf{I}}}\big(t_{\mathbf{I}}-\cdot\big)\big|f_{\mathbf{I}}\big(t_{\mathbf{I}}\big)\mathrm{d}t_{\mathbf{I}}\right\|_{\mathbf{I},\infty}$$

Put also $\xi_{h_{\mathbf{I}}}(\cdot)=\widetilde{f}_{h_{\mathbf{I}}}(\cdot)-s_{h_{\mathbf{I}}}(\cdot)$ and let

$$\zeta(h,\mathcal{P})=\sup_{\mathbf{I}\in\mathcal{P}}\|\xi_{h_{\mathbf{I}}}\|_{\mathbf{I},\infty},\quad \zeta_n=\sup_{\eta\in\mathcal{H}_n}\sup_{\mathcal{P}\in\mathfrak{P}}\Big[\zeta\big(\eta,\mathcal{P}\big)-\Lambda A_n(\eta,\mathcal{P})\Big]_+,$$

Lemma 2. *For any $p\ge1$ there exist $\mathbf{c_i}\big(p,d,\mathbf{K},\mathbf{f}\big)$, $\mathbf{i}=1,2,3,4$, such that for any $n\ge3$*

$$\textbf{(i)}\qquad \sup_{f\in\mathbf{F}(\mathbf{f})}\Big[\mathbb{E}^{(n)}_f\big(\zeta_n\big)^{2q}\Big]^{\frac{1}{2q}}\le \mathbf{c_1}\big(2q,d,\mathbf{K},\mathbf{f}\big)n^{-1/2};$$

$$\textbf{(ii)}\qquad \sup_{f\in\mathbf{F}(\mathbf{f})}\Big[\mathbb{E}^{(n)}_f\left[\bar{s}_n-\bar{\mathbf{f}}_n\right]_+^{2q}\Big]^{\frac{1}{2q}}\le \mathbf{c_2}\big(2q,d,\mathbf{K},\mathbf{f}\big)n^{-1/2};$$

$$\textbf{(iii)}\qquad \sup_{f\in\mathbf{F}(\mathbf{f})}\Big[\mathbb{E}^{(n)}_f\left[\bar{\mathbf{f}}_n-3\bar{s}_n\right]_+^{2q}\Big]^{\frac{1}{2q}}\le \mathbf{c_3}\big(2q,d,\mathbf{K},\mathbf{f}\big)n^{-1/2};$$

$$\textbf{(iv)}\qquad \sup_{f\in\mathbf{F}(\mathbf{f})}\Big[\mathbb{E}^{(n)}_f\big(\bar{\mathfrak{f}}_n\big)^{p}\Big]^{\frac{1}{p}}\le \mathbf{c_4}\big(p,d,\mathbf{K},\mathbf{f}\big).$$

The explicit expression of $\mathbf{c_i}\big(p,d,\mathbf{K},\mathbf{f}\big)$, $\mathbf{i}=1,2,3,4$ can be found in the proof of the lemma.

4.2.2. *Proof of Theorem 1.* We brake the proof on several steps.

$\mathbf{1^0}$. Let $\boldsymbol{h}\in\mathcal{H}_n$ and $\boldsymbol{\mathcal{P}}\in\mathfrak{P}$ be fixed. We have in view of triangle inequality

$$\Big\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}-f\Big\|_\infty\le\Big\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}-\widehat{f}_{(\boldsymbol{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})}\Big\|_\infty+\Big\|\widehat{f}_{(\boldsymbol{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})}-\widehat{f}_{\boldsymbol{h},\boldsymbol{\mathcal{P}}}\Big\|_\infty+\Big\|\widehat{f}_{\boldsymbol{h},\boldsymbol{\mathcal{P}}}-f\Big\|_\infty. \tag{4.1}$$

We have

$$\Big\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}-\widehat{f}_{(\boldsymbol{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})}\Big\|_\infty\le\widehat{\Delta}_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})+\lambda\widehat{A}_n\big(\widehat{h},\widehat{\mathcal{P}}\big). \tag{4.2}$$

Noting that $\widehat{f}_{(\boldsymbol{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})}\equiv\widehat{f}_{(\widehat{h},\widehat{\mathcal{P}})(\boldsymbol{h},\boldsymbol{\mathcal{P}})}$ we get

$$\Big\|\widehat{f}_{(\boldsymbol{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})}-\widehat{f}_{\boldsymbol{h},\boldsymbol{\mathcal{P}}}\Big\|_\infty\le\widehat{\Delta}_n\big(\widehat{h},\widehat{\mathcal{P}}\big)+\lambda\widehat{A}_n(\boldsymbol{h},\boldsymbol{\mathcal{P}}). \tag{4.3}$$

We obtain from (4.2) and (4.3)

$$\left\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}-\widehat{f}_{(\boldsymbol{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})}\right\|_\infty+\left\|\widehat{f}_{(\boldsymbol{h},\boldsymbol{\mathcal{P}})(\widehat{h},\widehat{\mathcal{P}})}-\widehat{f}_{\boldsymbol{h},\boldsymbol{\mathcal{P}}}\right\|_\infty$$
$$\leq\left[\widehat{\Delta}_n\big(\widehat{h},\widehat{\mathcal{P}}\big)+\lambda\widehat{A}_n\big(\widehat{h},\widehat{\mathcal{P}}\big)\right]+\left[\widehat{\Delta}_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})+\lambda\widehat{A}_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})\right]\leq 2\left[\widehat{\Delta}_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})+\lambda\widehat{A}_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})\right].$$

To get the last inequality we have used the definition of $(\widehat{h},\widehat{\mathcal{P}})$. Thus, we obtain from (4.1) that

$$\left\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}-f\right\|_\infty\leq 2\left[\widehat{\Delta}_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})+\lambda\widehat{A}_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})\right]+\left\|\widehat{f}_{\boldsymbol{h},\boldsymbol{\mathcal{P}}}-f\right\|_\infty. \tag{4.4}$$

2^{0}. Note that for any $h,\eta\in\mathcal{H}_n$ and any $\mathcal{P}'\in\mathfrak{P}$

$$\left\|\widehat{f}_{(h,\boldsymbol{\mathcal{P}}),(\eta,\mathcal{P}')}-\widehat{f}_{\eta,\mathcal{P}'}\right\|_\infty\leq d\big(\bar{\mathbf{f}}_n\big)^{\lfloor d^2/4\rfloor+1}\sup_{\mathbf{I}'\in\mathcal{P}'}\left\|\prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}:\ \boldsymbol{I}\cap\mathbf{I}'\neq\emptyset}\widetilde{f}_{h_{\boldsymbol{I}\cap\mathbf{I}'},\eta_{\boldsymbol{I}\cap\mathbf{I}'}}-\widetilde{f}_{\eta_{\mathbf{I}'}}\right\|_{\mathbf{I}',\infty}. \tag{4.5}$$

Here we have used the trivial inequality: for any $m\in\mathbb{N}^*$ and any $a_j,b_j:\mathcal{X}_j\to\mathbb{R},\ j=\overline{1,m}$,

$$\left\|\prod_{j=1}^m a_j-\prod_{j=1}^m b_j\right\|_\infty\leq m\Big(\sup_{j=\overline{1,m}}\|a_j-b_j\|_{\mathcal{X}_j,\infty}\Big)\Big[\sup_{j=\overline{1,m}}\max\left(\|a_j\|_{\mathcal{X}_j,\infty},\|b_j\|_{\mathcal{X}_j,\infty}\right)\Big]^{m-1}, \tag{4.6}$$

where $\|\cdot\|_{\mathcal{X}_j,\infty}$ and $\|\cdot\|_\infty$ denote the supremum norms on $\mathcal{X}_j$ and $\mathcal{X}_1\times\cdots\times\mathcal{X}_m$ respectively.

Introduce the following notation: for any $h,\eta\in\mathcal{H}_n$ and any $\mathcal{P}\in\mathfrak{P}$ we set

$$\xi^*_{h_\mathbf{I},\eta_\mathbf{I}}(\cdot)=\widetilde{f}_{h_\mathbf{I},\eta_\mathbf{I}}(\cdot)-s^*_{h_\mathbf{I},\eta_\mathbf{I}}(\cdot)$$

We have in view of (4.6) (here and later the product and the supremum over empty set are assumed equal to one and to zero respectively)

$$\left\|\prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}}\widetilde{f}_{h_{\boldsymbol{I}\cap\mathbf{I}'},\eta_{\boldsymbol{I}\cap\mathbf{I}'}}-\prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}}s^*_{h_{\boldsymbol{I}\cap\mathbf{I}'},\eta_{\boldsymbol{I}\cap\mathbf{I}'}}\right\|_{\mathbf{I}',\infty}\leq d\left[\max\left\{\bar{\mathbf{f}}_n,\mathrm{k}_1^2\mathbf{f}\right\}\right]^{d-1}\sup_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}}\left\|\xi^*_{h_{\boldsymbol{I}\cap\mathbf{I}'},\eta_{\boldsymbol{I}\cap\mathbf{I}'}}\right\|_{\boldsymbol{I}\cap\mathbf{I}',\infty}. \tag{4.7}$$

We remark that for any $\mathbf{I}\in\mathcal{I}_d$, any $h,\eta\in(0,1]^d$ and any $z_\mathbf{I}\in\mathbb{R}^{|\mathbf{I}|}$

$$\xi^*_{h_\mathbf{I},\eta_\mathbf{I}}(z_\mathbf{I})=\int_{\mathbb{R}^{|\mathbf{I}|}}K_{\eta_i}(z_\mathbf{I}-u_\mathbf{I})\xi_{h_\mathbf{I}}(u_\mathbf{I})\mathrm{d}u_\mathbf{I}$$

and, therefore,

$$\left\|\xi^*_{h_\mathbf{I},\eta_\mathbf{I}}\right\|_{\mathbf{I},\infty}\leq \mathrm{k}_1{}^{|\mathbf{I}|}\left\|\xi_{h_\mathbf{I}}\right\|_{\mathbf{I},\infty}\leq\mathrm{k}_1{}^d\left\|\xi_{h_\mathbf{I}}\right\|_{\mathbf{I},\infty},$$

since $\mathrm{k}_1\geq 1$ in view of Assumption 1. It yields together with (4.7)

$$\left\|\prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}}\widetilde{f}_{h_{\boldsymbol{I}\cap\mathbf{I}'},\eta_{\boldsymbol{I}\cap\mathbf{I}'}}-\prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}}s^*_{h_{\boldsymbol{I}\cap\mathbf{I}'},\eta_{\boldsymbol{I}\cap\mathbf{I}'}}\right\|_{\mathbf{I}',\infty}\leq d\mathrm{k}_1{}^d\left[\max\left\{\bar{\mathbf{f}}_n,\mathrm{k}_1^2\mathbf{f}\right\}\right]^{d-1}\sup_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}}\left\|\xi_{h_{\mathbf{I}\cap\mathbf{I}'}}\right\|_{\boldsymbol{I}\cap\mathbf{I}',\infty}. \tag{4.8}$$

Note also that for any $\eta\in\mathcal{H}_n$ and $\mathbf{I}'\in\mathcal{I}_d$

$$s_{\eta_{\mathbf{I}'}}(\cdot)=\int_{\mathbb{R}^{\mathbf{I}'}}K_{\eta_{\mathbf{I}'}}(t_{\mathbf{I}'}-\cdot)f_{\mathbf{I}'}(t_{\mathbf{I}'})\mathrm{d}t_{\mathbf{I}'}=\int_{\mathbb{R}^{\mathbf{I}'}}K_{\eta_{\mathbf{I}'}}(t_{\mathbf{I}'}-\cdot)\Big[\prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}}f_{\boldsymbol{I}\cap\mathbf{I}'}(t_{\boldsymbol{I}\cap\mathbf{I}'})\Big]\mathrm{d}t_{\mathbf{I}'}=\prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}}s_{\eta_{\boldsymbol{I}\cap\mathbf{I}'}}(\cdot).$$

Here we have used that $\boldsymbol{\mathcal{P}} \in \mathfrak{P}(f)$. Using once again (4.6) we obtain

$$\Big\| \prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}} s^*_{h_{\boldsymbol{I}\cap\mathbf{I}'},\eta_{\boldsymbol{I}\cap\mathbf{I}'}} - \prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}} s_{\eta_{\boldsymbol{I}\cap\mathbf{I}'}} \Big\|_{\mathbf{I}',\infty} \le d\left[\mathrm{k}_1^2\mathbf{f}\right]^{d-1} \sup_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}} \Big\| s^*_{h_{\boldsymbol{I}\cap\mathbf{I}'},\eta_{\boldsymbol{I}\cap\mathbf{I}'}} - s_{\eta_{\boldsymbol{I}\cap\mathbf{I}'}} \Big\|_{\boldsymbol{I}\cap\mathbf{I}',\infty}.$$

and, therefore, in view of Lemma 1

$$\Big\| \prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}} s^*_{h_{\boldsymbol{I}\cap\mathbf{I}'},\eta_{\boldsymbol{I}\cap\mathbf{I}'}} - s_{\eta_{\mathbf{I}'}} \Big\|_{\mathbf{I}',\infty} \quad\le\quad d\mathrm{k}_1^d\left[\mathrm{k}_1^2\mathbf{f}\right]^{d-1} \sup_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}} \big\| b_{h_{\boldsymbol{I}\cap\mathbf{I}'}} \big\|_{\boldsymbol{I}\cap\mathbf{I}',\infty}. \tag{4.9}$$

Thus, we obtain from (4.8) and (4.9)

$$\Big\| \prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}:\, \boldsymbol{I}\cap\mathbf{I}'\neq\emptyset} \widetilde{f}_{h_{\boldsymbol{I}\cap\mathbf{I}'},\eta_{\boldsymbol{I}\cap\mathbf{I}'}} - \widetilde{f}_{\eta_{\mathbf{I}'}} \Big\|_{\mathbf{I}',\infty} \le \mathfrak{f}_n\Big[\sup_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}} \big\|\xi_{h_{\boldsymbol{I}\cap\mathbf{I}'}}\big\|_{\boldsymbol{I}\cap\mathbf{I}',\infty} + \sup_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}} \big\|b_{h_{\boldsymbol{I}\cap\mathbf{I}'}}\big\|_{\boldsymbol{I}\cap\mathbf{I}',\infty}\Big] + \big\|\xi_{\eta_{\mathbf{I}'}}\big\|_{\mathbf{I}',\infty},$$

where we have put $\mathfrak{f}_n = 2d\mathrm{k}_1^d\left[\max\left\{\bar{\mathbf{f}}_n, \mathrm{k}_1^2\mathbf{f}\right\}\right]^{d-1}$.

Therefore, we get from (4.5) for any $h, \eta \in \mathcal{H}_n$ and $\mathcal{P}' \in \mathfrak{P}$

$$\Big\| \widehat{f}_{(h,\boldsymbol{\mathcal{P}}),(\eta,\mathcal{P}')} - \widehat{f}_{\eta,\mathcal{P}'} \Big\|_\infty \le \bar{\mathfrak{f}}_n \Big\{ \zeta\big(h, \boldsymbol{\mathcal{P}} \diamond \mathcal{P}'\big) + \sup_{\mathbf{I}\in\boldsymbol{\mathcal{P}}\diamond\mathcal{P}'} \|b_{h_{\mathbf{I}}}\|_{\mathbf{I},\infty} \Big\} + \tilde{\mathfrak{f}}_n \zeta\big(\eta, \mathcal{P}'\big). \tag{4.10}$$

Here we have put $\tilde{\mathfrak{f}}_n = d\big(\bar{\mathbf{f}}_n\big)^{\lfloor d^2/4\rfloor+1}$ and $\bar{\mathfrak{f}}_n = \tilde{\mathfrak{f}}_n\mathfrak{f}_n$ Taking into account that for any $h \in \mathcal{H}_n$ and any $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$A_n\big(h, \mathcal{P} \diamond \mathcal{P}'\big) \le A_n\big(h, \mathcal{P}\big) \wedge A_n\big(h, \mathcal{P}'\big),$$

we get from (4.10)

$$\Big\| \widehat{f}_{(h,\boldsymbol{\mathcal{P}}),(\eta,\mathcal{P}')} - \widehat{f}_{\eta,\mathcal{P}'} \Big\|_\infty \le \bar{\mathfrak{f}}_n \Big\{ \Lambda A_n\big(h, \boldsymbol{\mathcal{P}}\big) + \sup_{\mathbf{I}\in\boldsymbol{\mathcal{P}}\diamond\mathcal{P}'} \|b_{h_{\mathbf{I}}}\|_{\mathbf{I},\infty} + \zeta_n \Big\} + \tilde{\mathfrak{f}}_n \zeta\big(\eta, \mathcal{P}'\big). \tag{4.11}$$

Remembering that $\lambda = \tilde{\mathfrak{f}}_n\Lambda$, we obtain from (4.11)

$$\widehat{\Delta}_n(\boldsymbol{h}, \boldsymbol{\mathcal{P}}) \le \bar{\mathfrak{f}}_n \Big\{ \Lambda A_n\big(\boldsymbol{h}, \boldsymbol{\mathcal{P}}\big) + B\big(\boldsymbol{h}, \boldsymbol{\mathcal{P}}\big) + \zeta_n \Big\} + \tilde{\mathfrak{f}}_n \Big\{ \zeta_n + \Lambda \sup_{\eta\in\mathcal{H}_n} \sup_{\mathcal{P}\in\mathcal{P}} \Big[A_n(\eta, \mathcal{P}) - \widehat{A}_n(\eta, \mathcal{P}) \Big]_+ \Big\},$$

where, remind $B\big(h, \mathcal{P}\big) = \sup_{\mathcal{P}'} \sup_{\mathbf{I}\in\mathcal{P}\diamond\mathcal{P}'} \|b_{h_{\mathbf{I}}}\|_{\mathbf{I},\infty}$.

Taking into account that $\bar{\mathfrak{f}}_n \ge \tilde{\mathfrak{f}}_n$, since $\bar{\mathfrak{f}}_n \ge 1$ we finally get

$$\widehat{\Delta}_n(\boldsymbol{h}, \boldsymbol{\mathcal{P}}) \le \bar{\mathfrak{f}}_n \Big\{ \Lambda A_n\big(\boldsymbol{h}, \boldsymbol{\mathcal{P}}\big) + B\big(\boldsymbol{h}, \boldsymbol{\mathcal{P}}\big) + 2\zeta_n + \Lambda \sup_{\eta\in\mathcal{H}_n} \sup_{\mathcal{P}\in\mathcal{P}} \Big[A_n(\eta, \mathcal{P}) - \widehat{A}_n(\eta, \mathcal{P}) \Big]_+ \Big\}. \tag{4.12}$$

Note that the definition of $\mathcal{H}_n$ implies that

$$\Big[A_n(\eta, \mathcal{P}) - \widehat{A}_n(\eta, \mathcal{P}) \Big]_+ \le \mathfrak{a}^* \left[\sqrt{\bar{s}_n} - \sqrt{\bar{\mathbf{f}}_n} \right]_+ \le \mathfrak{a}^* \left[\bar{s}_n - \bar{\mathbf{f}}_n \right]_+, \quad \forall \eta \in \mathcal{H}_n,\ \forall \mathcal{P} \in \mathcal{P}.$$

To get the last inequality we have also used that by definition $\bar{\mathbf{f}}_n, \bar{s}_n \ge 1$.

Putting $R_n = \mathfrak{a}^*\Lambda\left[\bar{s}_n - \bar{\mathbf{f}}_n\right]_+$ we obtain in view of (4.12)

$$\widehat{\Delta}_n(\boldsymbol{h}, \boldsymbol{\mathcal{P}}) \le \bar{\mathfrak{f}}_n \Big\{ \Lambda A_n\big(\boldsymbol{h}, \boldsymbol{\mathcal{P}}\big) + B\big(\boldsymbol{h}, \boldsymbol{\mathcal{P}}\big) + 2\zeta_n + R_n \Big\}, \tag{4.13}$$

Note also that the definition of $\mathcal{H}_n$ implies that

$$\left[\widehat{A}_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})-\sqrt{3}A_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})\right]_+\leq \mathfrak{a}^*\left[\sqrt{\bar{\mathbf{f}}_n}-\sqrt{3\bar{s}_n}\right]_+\leq \mathfrak{a}^*\left[\bar{\mathbf{f}}_n-3\bar{s}_n\right]_+,\quad \forall\eta\in\mathcal{H}_n,\ \forall\mathcal{P}\in\mathcal{P}.$$

Thus, denoting $\mathcal{R}_n=\mathfrak{a}^*\Lambda\left[\bar{\mathbf{f}}_n-3\bar{s}_n\right]_+$ we obtain using (4.13)

$$\widehat{\Delta}_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})+\lambda\widehat{A}_n\big(\boldsymbol{h},\boldsymbol{\mathcal{P}}\big)\leq\bar{\mathfrak{f}}_n\Big\{3\Lambda A_n\big(\boldsymbol{h},\boldsymbol{\mathcal{P}}\big)+B\big(\boldsymbol{h},\boldsymbol{\mathcal{P}}\big)+2\zeta_n+R_n+\mathcal{R}_n\Big\},\tag{4.14}$$

where we have used also $\sqrt{3}<2$.

3^0. Note that in view of $\boldsymbol{\mathcal{P}}\in\mathfrak{P}(f)$, (4.6) and (4.13)

$$\begin{aligned}\left\|\widehat{f}_{\boldsymbol{h},\boldsymbol{\mathcal{P}}}-f\right\|_\infty &= \left\|\prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}}\widetilde{f}_{\boldsymbol{h}_{\boldsymbol{I}}}(x_{\boldsymbol{I}})-\prod_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}}f_{\boldsymbol{I}}(x_{\boldsymbol{I}})\right\|_\infty\\ &\leq d\left[\max\left\{\bar{\mathbf{f}}_n,\mathrm{k}_1^2\mathbf{f}\right\}\right]^{d-1}\sup_{\boldsymbol{I}\in\boldsymbol{\mathcal{P}}}\left\|\widetilde{f}_{\boldsymbol{h}_{\boldsymbol{I}}}(x_{\boldsymbol{I}})-f_{\boldsymbol{I}}(x_{\boldsymbol{I}})\right\|_{\boldsymbol{I},\infty}\\ &\leq d\left[\max\left\{\bar{\mathbf{f}}_n,\mathrm{k}_1^2\mathbf{f}\right\}\right]^{d-1}\Big[B(\boldsymbol{h},\boldsymbol{\mathcal{P}})+\zeta(\boldsymbol{h},\boldsymbol{\mathcal{P}})\Big]\leq\bar{\mathfrak{f}}_n\Big[B(\boldsymbol{h},\boldsymbol{\mathcal{P}})+\Lambda A_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})+\zeta_n\Big].\end{aligned}\tag{4.15}$$

Here we have also used that $\mathcal{P}\equiv\mathcal{P}\diamond\mathcal{P}$. We obtain from (4.4), (4.14) and (4.15)

$$\left\|\widehat{f}_{\widehat{h},\widehat{\mathcal{P}}}-f\right\|_\infty\leq\bar{\mathfrak{f}}_n\left[3B\big(\boldsymbol{h},\boldsymbol{\mathcal{P}}\big)+7\Lambda A_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})+5\zeta_n+2R_n+2\mathcal{R}_n\right],$$

and, therefore, for any $\boldsymbol{h}\in\mathcal{H}_n$, $\boldsymbol{\mathcal{P}}\in\mathfrak{P}$ and $q\geq1$

$$\Big(\mathbb{E}_f^{(n)}\|\widehat{f}_{\widehat{h}[\widehat{\mathcal{P}}],\widehat{\mathcal{P}}}-f\|_\infty\Big)^{\frac{1}{q}}\leq E_q\Big[3B\big(\boldsymbol{h},\boldsymbol{\mathcal{P}}\big)+7\Lambda A_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})\Big]+E_{2q}\Big[5y_{1,n}+2\Lambda\mathfrak{a}^*\big(y_{2,n}+y_{3,n}\big)\Big],\tag{4.16}$$

where we have put for $p\geq1$

$$E_p=\left[\mathbb{E}_f^{(n)}\big(\bar{\mathfrak{f}}_n\big)^p\right]^{\frac{1}{p}},\,y_{1,n}=\left[\mathbb{E}_f^{(n)}\big(\zeta_n\big)^{2q}\right]^{\frac{1}{2q}},\,y_{2,n}=\left[\mathbb{E}_f^{(n)}\left[\bar{s}_n-\bar{\mathbf{f}}_n\right]_+^{2q}\right]^{\frac{1}{2q}},\,y_{3,n}=\left[\mathbb{E}_f^{(n)}\left[\bar{\mathbf{f}}_n-3\bar{s}_n\right]_+^{2q}\right]^{\frac{1}{2q}}.$$

Taking into account that the right hand side of (4.16) is independent of the choice $\boldsymbol{h}$ and $\boldsymbol{\mathcal{P}}$ and that the quantity $\bar{s}_n\leq1\vee[\mathrm{k}_1\mathbf{f}]$ we get

$$\begin{aligned}\Big(\mathbb{E}_f^{(n)}\|\widehat{f}_{\widehat{h}[\widehat{\mathcal{P}}],\widehat{\mathcal{P}}}-f\|_\infty\Big)^{\frac{1}{q}} &\leq 7\Lambda E_q\bigg(\inf_{\boldsymbol{h}\in\mathcal{H}_n}\inf_{\boldsymbol{\mathcal{P}}\in\mathfrak{P}(f)}\Big[B\big(\boldsymbol{h},\boldsymbol{\mathcal{P}}\big)+A_n(\boldsymbol{h},\boldsymbol{\mathcal{P}})\Big]\bigg)\\ &\quad+E_{2q}\Big[5y_{1,n}+2\Lambda\mathfrak{a}^*\big(y_{2,n}+y_{3,n}\big)\Big]\\ &= \mathbf{C_1}\big(q,d,\mathbf{K},\mathbf{f}\big)\mathfrak{R}(f)+E_{2q}\Big[5y_{1,n}+2\Lambda\mathfrak{a}^*\big(y_{2,n}+y_{3,n}\big)\Big].\end{aligned}$$

where we have put $\mathbf{C_1}\big(q,d,\mathbf{K},\mathbf{f}\big)=7\Lambda E_q\sqrt{1\vee[\mathrm{k}_1\mathbf{f}]}$.

This inequality together with bounds found in Lemma 2 leads to the assertion of the theorem.

∎

4.3. *Proof of Theorem 2.* The proof of Theorem 2 is relatively standard and based on the general result established in Kerkyacharian, Lepski and Picard (2007), Proposition 7. For the convenience we formulate this result not in full generality but its version reduced to the considered problem. Let $\big(\beta,p,\mathcal{P}\big)\in(0,\infty)^d\times[1,\infty]^d\times\mathfrak{P}$ such that $\Upsilon\big(\beta,p,\mathcal{P}\big)>0$ and $\mathcal{L}\in(0,\infty)^d$ be fixed.

LEMMA 3. *Assume that there exist $f_0\in\mathbf{N}_{p,d}\big(\beta,\mathcal{L},\mathcal{P}\big)$, $\rho_n>0,\ n\in\mathbb{N}^*$, and a finite set $\mathbf{J}_n$ such that for any sufficiently large $n\in\mathbb{N}^*$ one can find $\big\{f^{(\mathbf{j})},\ \mathbf{j}\in\mathbf{J}_n\big\}\subset\mathbf{N}_{p,d}\big(\beta,\mathcal{L},\mathcal{P}\big)$ satisfying*

$$\|f^{(\mathbf{j})}-f_0\|_\infty=\rho_n,\qquad\forall\mathbf{j}\in\mathbf{J}_n;\tag{4.17}$$

$$\limsup_{n\to\infty}\mathbb{E}^{(n)}_{f_0}\left[\frac{1}{|\mathbf{J}_n|}\sum_{\mathbf{j}\in\mathbf{J}_n}\frac{\mathrm{d}\mathbb{P}^{(n)}_{f^{(\mathbf{j})}}}{\mathrm{d}\mathbb{P}^{(n)}_{f_0}}\Big(X^{(n)}\Big)-1\right]^2=:\mathbf{C}<\infty.\tag{4.18}$$

Then for $r\geq1$

$$\liminf_{n\to\infty}\inf_{\tilde f}\sup_{f\in\mathbf{N}_{p,d}\left(\beta,\mathcal{L},\mathcal{P}\right)}\rho_n^{-1}\left(\mathbb{E}^{(n)}_f\big\|\tilde f-f\big\|^r_\infty\right)^{\frac1r}\geq2^{-1}\left[1-\sqrt{\mathbf{C}/(\mathbf{C}+4)}\right],$$

where infimum is taken over all possible estimators.

Proof of the theorem. Set $\mathcal{N}(x)=\prod_{i=1}^d\bigg([2\pi]^{-1/2}\exp-\big\{x_i^2/2\big\}\bigg)$ and let $f_0(x)=\sigma^{-1}\mathcal{N}(x/\sigma)$, where $\sigma>0$ is chosen in such a way that f_0 belongs to the class $\overline{\mathbb{N}}_{p,d}\big(\beta,\mathcal{L}/2\big)$. We remark that in order to obey the latter restriction it suffices to choose σ satisfying

$$\sup_{\mathbf{I}\in\mathcal{I}_d}\sup_{i\in\mathbf{I}}\sigma^{-\beta_i+|\mathbf{I}|/r_i}\big\|\mathcal{N}_\mathbf{I}\big\|^{|\mathbf{I}|}_{r_i}\leq\inf_{i=\overline{1,d}}\mathcal{L}_i.\tag{4.19}$$

The product structure of f_0 together with (4.19) allows us to assert that $f_0\in\mathbf{N}_{p,d}\big(\beta,\mathcal{L}/2,\mathcal{P}\big)$ for any $\mathcal{P}\in\mathfrak{P}$. Let $\mathbf{I}^*\in\{1,\dots,d\}$ be defined from the relation

$$\Upsilon\big(\beta,p,\mathcal{P}\big):=\inf_{\mathbf{I}\in\mathcal{P}}\gamma_\mathbf{I}(\beta,p)=\gamma_{\mathbf{I}^*}(\beta,p),$$

and for the notation convenience the elements of $\mathbf{I}^*$ will be denoted by $i_1,\dots,i_m$ and $m=|I^*|$.

Let $g:\mathbb{R}\to\mathbb{R}$ be compactly supported on $(-1/2,1/2)$ function, satisfying $g\in\cap_{i\in\mathbf{I}^*}N_{p_i,1}(\beta_i,1/2)$, and such that $\int g=0$. Suppose also that $\big|g(0)\big|=\|g\|_\infty$.

Let $A_n\to0$ and $\delta_{l,n}\to0,\ l=\overline{1,m},\ n\to\infty$, be sequences whose choice will be done later and set $\mathbf{J}_n:=\big[1,\dots,M_{1,n}\big]\times\cdots\times\big[1,\dots,M_{m,n}\big]\subset\mathbb{N}^m$, where $M_{l,n}=\big\lfloor\delta_{l,n}^{-1/2}\big\rfloor,\ l=\overline{1,m}$.

For any $\mathbf{j}=\big(j_1,\dots j_m\big)\in\mathbf{J}_n$ define $G_\mathbf{j}(x_\mathbf{I})=A_n\prod_{l=1}^m g\left(\delta^{-1}_{i,n}\Big[x_{i_l}-x^{(\mathbf{j})}_{i_l}\Big]\right)$. Here for any $\mathbf{j}\in\mathbf{J}_n$ we put $x^{(\mathbf{j})}_{i_l}=j_l\delta_{l,n}$. The choice of g implies

$$G_\mathbf{j}G_{\mathbf{j}'}\equiv0,\quad\forall\mathbf{j},\mathbf{j}'\in\mathbf{J}_n,\ \mathbf{j}\neq\mathbf{j}'.\tag{4.20}$$

Note also that the system of equations

$$A_n\delta^{-\beta_{i_k}}_{k,n}\left(\prod_{l=1}^m\delta_{l,n}\right)^{1/p_{i_k}}=\frac{\mathcal{L}_{i_k}}{c_k},\quad k=\overline{1,m},\tag{4.21}$$

implies that $G_{\mathbf{j}} \in N_{p_{\mathbf{I}},d}\big(\beta_{\mathbf{I}}, \mathcal{L}_{\mathbf{I}}/2\big)$ for any $\mathbf{j} \in \mathbf{J}_n$. Here we have denoted $c_k = \big(\|g\|_{p_{i_k}}\big)^{m-1}$.

Introduce the family of functions $\big\{f^{(\mathbf{j})},\ \mathbf{j} \in \mathbf{J}_n\big\}$ as follows.

$$f^{(\mathbf{j})}(x) = \prod_{i \notin \mathbf{I}^*}^{d} \bigg(\big[2\pi\sigma^2\big]^{-1/2} \exp -\big\{x_i^2/2\sigma^2\big\}\bigg)\bigg(\prod_{i \in \mathbf{I}^*}^{d} \big[2\pi\sigma^2\big]^{-1/2} \exp -\big\{x_i^2/2\sigma^2\big\} + G_{\mathbf{j}}\big(x_{\mathbf{I}}\big)\bigg).$$

First we remark that $A_n \to 0,\ n \to \infty$, implies that $f^{(\mathbf{j})} > 0$ for all sufficiently large n. Next, the assumption $\int g = 0$ implies that $\int f^{(\mathbf{j})} = 1$. Thus, $f^{(\mathbf{j})}$ is a probability density for any $\mathbf{j} \in \mathbf{J}_n$ for all sufficiently large n. At last the choice of f_0 together with (4.21) allows us to assert that $f^{(\mathbf{j})} \in \mathbf{N}_{p,d}\big(\beta, \mathcal{L}, \mathcal{P}\big)$ for any $\mathbf{j} \in \mathbf{J}_n$.

Thus, we conclude that Lemma 3 is applicable to the family $\big\{f^{(\mathbf{j})},\ \mathbf{j} \in \mathbf{J}_n\big\}$. We remark also that

$$\big\|f^{(\mathbf{j})} - f_0\big\|_\infty = c_1^* A_n, \quad \forall \mathbf{j} \in \mathbf{J}_n, \tag{4.22}$$

where we have put $c_1^* = |g(0)|^m \Big(2\pi\sigma^2\Big)^{(m-d)/2}$. Here we have also used that $\big|g(0)\big| = \|g\|_\infty$. We conclude that the assumption (4.17) is fulfilled with $\rho_n = c_1^* A_n$.

Let us now proceed with the verification of the condition (4.18) of Lemma 3. Note first that

$$\frac{\mathrm{d}\mathbb{P}^{(n)}_{f^{(\mathbf{j})}}}{\mathrm{d}\mathbb{P}^{(n)}_{f_0}}\Big(X^{(n)}\Big) = \prod_{k=1}^{n} \frac{f^{(\mathbf{j})}(X_k)}{f_0(X_k)}$$

and, therefore,

$$\bigg[\frac{1}{|\mathbf{J}_n|} \sum_{\mathbf{j} \in \mathbf{J}_n} \frac{\mathrm{d}\mathbb{P}^{(n)}_{f^{(\mathbf{j})}}}{\mathrm{d}\mathbb{P}^{(n)}_{f_0}}\Big(X^{(n)}\Big)\bigg]^2 = \frac{1}{|\mathbf{J}_n|^2}\bigg\{ \sum_{\mathbf{j} \in \mathbf{J}_n} \prod_{k=1}^{n} \bigg[\frac{f^{(\mathbf{j})}(X_k)}{f_0(X_k)}\bigg]^2 + \sum_{\substack{\mathbf{j},\mathbf{j}' \in \mathbf{J}_n:\\ \mathbf{j} \neq \mathbf{j}'}} \prod_{k=1}^{n} \frac{f^{(\mathbf{j})}(X_k) f^{(\mathbf{j}')}(X_k)}{f_0^2(X_k)} \bigg\}. \tag{4.23}$$

Since $X_k,\ k = \overline{1,n}$ are i.i.d. random vectors, we have for any $\mathbf{j} \neq \mathbf{j}'$

$$\mathbb{E}^{(n)}_{f_0}\bigg\{ \prod_{k=1}^{n} \frac{f^{(\mathbf{j})}(X_k) f^{(\mathbf{j}')}(X_k)}{f_0^2(X_k)}\bigg\} = \bigg\{ \int_{\mathbb{R}^{|\mathbf{I}^*|}} \bigg[1 + \frac{G_{\mathbf{j}}\big(x_{\mathbf{I}^*}\big)}{f_{\mathbf{I}^*,0}\big(x_{\mathbf{I}^*}\big)}\bigg]\bigg[1 + \frac{G_{\mathbf{j}'}\big(x_{\mathbf{I}^*}\big)}{f_{\mathbf{I}^*,0}\big(x_{\mathbf{I}^*}\big)}\bigg] f_{\mathbf{I}^*,0}\big(x_{\mathbf{I}^*}\big)\mathrm{d}x_{\mathbf{I}^*}\bigg\}^n = 1.$$

To get the last equality we have used (4.20) and the fact that $\int_{\mathbb{R}^{|\mathbf{I}^*|}} G_{\mathbf{j}}\big(x_{\mathbf{I}^*}\big)\mathrm{d}x_{\mathbf{I}^*} = 0$ since $\int g = 0$.

The latter result together with (4.23) yields

$$\begin{aligned}
\mathcal{E}_n \quad &:= \quad \mathbb{E}^{(n)}_{f_0}\bigg[\frac{1}{|\mathbf{J}_n|} \sum_{\mathbf{j} \in \mathbf{J}_n} \frac{\mathrm{d}\mathbb{P}^{(n)}_{f^{(\mathbf{j})}}}{\mathrm{d}\mathbb{P}^{(n)}_{f_0}}\Big(X^{(n)}\Big) - 1\bigg]^2 \\
&= \quad \frac{1}{|\mathbf{J}_n|^2} \sum_{\mathbf{j} \in \mathbf{J}_n} \bigg\{ \int_{\mathbb{R}^{|\mathbf{I}^*|}} \bigg[1 + \frac{G_{\mathbf{j}}\big(x_{\mathbf{I}^*}\big)}{f_{\mathbf{I}^*,0}\big(x_{\mathbf{I}^*}\big)}\bigg]^2 f_{\mathbf{I}^*,0}\big(x_{\mathbf{I}^*}\big)\mathrm{d}x_{\mathbf{I}^*}\bigg\}^n - |\mathbf{J}_n|^{-1} \\
&= \quad \frac{1}{|\mathbf{J}_n|^2} \sum_{\mathbf{j} \in \mathbf{J}_n} \bigg\{1 + \int_{\mathbb{R}^m} \bigg[\frac{G^2_{\mathbf{j}}(y)}{f_{\mathbf{I}^*,0}(y)}\bigg]\mathrm{d}y\bigg\}^n - |\mathbf{J}_n|^{-1}. \qquad (4.24)
\end{aligned}$$

Since, $G_{\mathbf{j}}(y) = 0$ for any $y \notin \left[0, \sqrt{\delta_{1,n}}\right] \times \cdot \times \left[0, \sqrt{\delta_{m,n}}\right] =: \mathcal{Y}_n$ we have for all n large enough $\inf_{y\in\mathcal{Y}_n} f_{\mathbf{I}^*,0}(y) \geq 2^{-1}\left(2\pi\sigma^2\right)^{-m}$. It yields together with (4.23), putting $c_2^* = 2\left(2\pi\sigma^2\right)^m \|g\|_2^{2m}$,

$$\mathcal{E}_n \leq |\mathbf{J}_n|^{-1}\left(1 + c_2^* A_n^2 \prod_{l=1}^m \delta_{l,n}\right)^n.$$

If we choose A_n and $\delta_{l,n},\ l = \overline{1,m}$ satisfying

$$c_2^* n A_n^2 \prod_{l=1}^m \delta_{l,n} \leq (1/4)\ln\left(\prod_{l=1}^m \delta_{l,n}^{-1}\right) \leq \ln\left(|\mathbf{J}_n|\right), \tag{4.25}$$

for all $n \geq 1$ large enough, then $\mathcal{E}_n \leq 1$ and, therefore, the condition (4.18) is fulfilled with $\mathbf{C} = 1$.

Thus, we have to choose A_n and $\delta_{l,n},\ l = \overline{1,m}$ satisfying (4.21) and (4.25). Let $t > 0$ be the number whose choice will be done later. Consider instead of (4.25) the equation

$$n A_n^2 \prod_{l=1}^m \delta_{l,n} = t^2 \ln(n). \tag{4.26}$$

and solve (4.21) and (4.26). Straightforward computations yield

$$A_n = R(\varepsilon t)^{\frac{1-\sum_{l=1}^m \frac{1}{\beta_{i_l} p_{i_l}}}{1-\sum_{l=1}^m \left(\frac{1}{p_{i_l}} - \frac{1}{2}\right)\frac{1}{\beta_{i_l}}}}, \qquad \delta_{l,n} = A_n^{\frac{1}{\beta_{i_l}} - \frac{2}{\beta_{i_l} p_{i_l}}} (t\varepsilon)^{\frac{2}{\beta_{i_l} p_{i_l}}} \left(c_l/\mathcal{L}_l\right)^{\frac{1}{\beta_{i_l}}},$$

where we have put $R = \left(\prod_{l=1}^m \left(c_l/\mathcal{L}_l\right)^{\frac{1}{2\beta_{i_l}}}\right)^{\frac{1}{1-\sum_{l=1}^m \left(\frac{1}{p_{i_l}} - \frac{1}{2}\right)\frac{1}{\beta_{i_l}}}}$. Moreover we have in view of (4.26)

$$\left(\prod_{l=1}^m \delta_{l,n}\right)^{-1/2} = R(\varepsilon t)^{-a}, \quad a = \frac{\sum_{l=1}^m \frac{1}{\beta_{i_l}}}{1 - \sum_{l=1}^m \left(\frac{1}{p_{i_l}} - \frac{1}{2}\right)\frac{1}{\beta_{i_l}}}$$

and, therefore, $(1/4)\ln\left(\prod_{l=1}^m \delta_{l,n}^{-1}\right) \asymp (a/2)\ln(n), \quad n \to \infty$, Hence, choosing t as an arbitrary number satisfying $t^2 < (2c_2^*)^{-1} a$ we guarantee that (4.26) implies (4.25) for all n large enough.

Thus, we conclude that Lemma 3 is applicable with

$$\rho_n = c_1^* A_n = c_1^* R\left(\frac{t\ln(n)}{n}\right)^{\frac{1-\sum_{l=1}^m \frac{1}{\beta_{i_l} p_{i_l}}}{2\left(1-\sum_{l=1}^m \left[\frac{1}{p_{i_l}} - \frac{1}{2}\right]\frac{1}{\beta_{i_l}}\right)}}.$$

It remains to note that the definition of I^* implies that $\Upsilon\left(\beta, p, \mathcal{P}\right) = \frac{1-\sum_{l=1}^m \frac{1}{\beta_{i_l} p_{i_l}}}{\sum_{l=1}^m \frac{1}{\beta_{i_l}}}$. We remark that

$$\frac{\Upsilon\left(\beta, p, \mathcal{P}\right)}{2\Upsilon\left(\beta, p, \mathcal{P}\right) + 1} = \frac{1 - \sum_{l=1}^m \frac{1}{\beta_{i_l} p_{i_l}}}{2\left(1 - \sum_{l=1}^m \left[\frac{1}{p_{i_l}} - \frac{1}{2}\right]\frac{1}{\beta_{i_l}}\right)}$$

and the assertion of the theorem follows. ∎

4.4. *Proof of Theorem 3.* The proof of the theorem is based on the application of Theorem 1 and on Lemma 4 below that allows us to bound from above the quantity $B(h,\mathcal{P})$. The assertion of the lemma, whose proof is postponed to Appendix, is based on the embedding theorem for anisotropic Nikolskii spaces. For any function $g:\mathbb{R}^s\to\mathbb{R}$ and any $\eta\in(0,\infty)^s$ set

$$\mathcal{B}_{\eta,g}(z)=\int_{\mathbb{R}^s}K_\eta(t-z)g(t)\mathrm{d}t-g(z),\quad z\in\mathbb{R}^s.$$

Lemma 4. *Let* $\mathbf{K}$ *satisfy Assumption 1 and (3.1). Let* $(\alpha,r)\in(0,\mathfrak{b}]^s\times[1,\infty]^s$ *be such that* $\varkappa=1-\sum_{l=1}^s(\alpha_l r_l)^{-1}>0$ *and let* $Q\in(0,\infty)^s$*. Then there exists* $\mathrm{c}=\mathrm{c}\big(s,r,\mathfrak{b}\big)>0$ *such that*

$$\sup_{g\in\mathbb{N}_{r,s}(\alpha,Q)}\|\mathcal{B}_{\eta,g}\|_\infty\leq \mathrm{c}\mathrm{k}_1^s\sum_{i=1}^s Q_i\eta_i^{\boldsymbol{\alpha}_i},\quad\forall\eta\in(0,\infty)^s.$$

Here $\boldsymbol{\alpha}=(\boldsymbol{\alpha}_1,\dots\boldsymbol{\alpha}_s)$, $\boldsymbol{\alpha}_i=\varkappa\alpha_i\varkappa_i^{-1}$ *and* $\varkappa_i=1-\sum_{l=1}^s\big(r_l^{-1}-r_i^{-1}\big)\alpha_l^{-1}$.

Proof of Theorem 3. Let $\big(\beta,p,\mathcal{P}\big)\in(0,\mathfrak{b}]^d\times[1,\infty]^d\times\mathfrak{P}$ such that $\Upsilon\big(\beta,p,\mathcal{P}\big)>0$ and $\mathcal{L}\in(0,\infty)^d$ be fixed. For any $\mathbf{I}\in\mathcal{I}_d$ and any $\mathbf{i}\in\mathbf{I}$ define

$$\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})=\tau(\mathbf{I})\beta_{\mathbf{i}}\tau_{\mathbf{i}}^{-1}(\mathbf{I}),\quad\tau(\mathbf{I})=1-\sum_{l\in\mathbf{I}}(\beta_l p_l)^{-1},\quad\tau_{\mathbf{i}}(\mathbf{I})=1-\sum_{l\in\mathbf{I}}\big(p_l^{-1}-p_{\mathbf{i}}^{-1}\big)\beta_l^{-1},$$

and remark that the condition $\Upsilon\big(\beta,p,\mathcal{P}\big)>0$ implies that $\tau(\mathbf{I})>0$ for any $\mathbf{I}\in\mathcal{I}_d$.

Let us first prove the following simple fact. Denote $\mathcal{C}_{\mathbf{i}}(\mathbf{I})=\{\mathbf{J}\subseteq\mathbf{I}:\ \mathbf{i}\in\mathbf{J}\}$, $\mathbf{i}\in\mathbf{I}$. Then

$$\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})=\inf_{\mathbf{J}\in\mathcal{C}_{\mathbf{i}}(\mathbf{I})}\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{J}),\quad\forall\mathbf{i}\in\mathbf{I}.\tag{4.27}$$

Indeed, we remark that $\tau_{\mathbf{i}}(\mathbf{J})=1-\sum_{l\in\mathbf{J}}\big(p_l^{-1}-p_{\mathbf{i}}^{-1}\big)\beta_l^{-1}=\tau(\mathbf{J})+p_{\mathbf{i}}^{-1}\sum_{l\in\mathbf{J}}\beta_l^{-1}$ and, therefore,

$$\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{J})=\frac{\beta_{\mathbf{i}}\tau(\mathbf{J})}{\tau(\mathbf{J})+p_{\mathbf{i}}^{-1}\beta^{-1}(\mathbf{J})},\qquad\beta^{-1}(\mathbf{J})=\sum_{l\in\mathbf{J}}\beta_l^{-1}.$$

We obviously have $\tau(\mathbf{J})\geq\tau(\mathbf{I})$ and $\beta^{-1}(\mathbf{J})\leq\beta^{-1}(\mathbf{I})$ for any $\mathbf{J}\subseteq\mathbf{I}$. It remains to note that $x\mapsto x/(x+a)$ is increasing on $\mathbb{R}_+$ for any $a>0$ and (4.27) follows.

Let $\mathcal{P}'\in\mathfrak{P}$ be an arbitrary partition. Since $f\in\overline{\mathbb{N}}_{p,d}\big(\beta,\mathcal{L}\big)$ we have $f_{\mathbf{J}}\in\mathbb{N}_{p_{\mathbf{J}},|\mathbf{J}|}\big(\beta_{\mathbf{J}},\mathcal{L}_{\mathbf{J}}\big)$ and, therefore, in view of Lemma 4 we have for any $h\in(0,1]^d$ and $\mathbf{J}\in\mathcal{P}\diamond\mathcal{P}'$

$$b_{h_{\mathbf{J}}}\leq\mathrm{c}\big(|\mathbf{J}|,p_{\mathbf{J}},\mathfrak{b}\big)\mathrm{k}_1^{|\mathbf{J}|}\sum_{\mathbf{i}\in\mathbf{J}}\mathcal{L}_{\mathbf{i}}h_{\mathbf{i}}^{\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{J})}\leq\mathrm{c}_1\sum_{\mathbf{i}\in\mathbf{I}}\mathcal{L}_{\mathbf{i}}h_{\mathbf{i}}^{\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})}.$$

To get the last inequality we use (4.27), $h\in(0,1]^d$ and we have put $\mathrm{c}_1=\mathrm{k}_1^d\sup_{\mathbf{J}\in\mathcal{I}_d}\mathrm{c}\big(|\mathbf{J}|,p_{\mathbf{J}},\mathfrak{b}\big)\mathrm{k}_1^{|\mathbf{J}|}$.

Noting that the right hand side of the latter inequality is independent on $\mathbf{J}$ we obtain

$$B\big(h,\mathcal{P}\big)\leq\mathrm{c}_1\sup_{\mathbf{I}\in\mathcal{P}}\sum_{\mathbf{i}\in\mathbf{I}}\mathcal{L}_{\mathbf{i}}h_{\mathbf{i}}^{\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})},\quad h\in(0,1]^d.$$

It remains to choose multi-bandwidth h. To do it it suffices to solve for any $\mathbf{I}\in\mathcal{P}$ the following system of equations.

$$\mathcal{L}_{\mathbf{j}}h_{\mathbf{j}}^{\boldsymbol{\beta}_{\mathbf{j}}(\mathbf{I})}=\mathcal{L}_{\mathbf{i}}h_{\mathbf{i}}^{\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})}=\sqrt{\frac{\ln(n)}{nV_{h_{\mathbf{I}}}}},\quad\mathbf{i},\mathbf{j}\in\mathbf{I}.$$

The solution is given by

$$h_{\mathbf{i}}=\mathcal{L}^{-\frac{1}{\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})}}\left(\frac{\mathfrak{L}(\mathbf{I})\ln(n)}{n}\right)^{\frac{\gamma_{\mathbf{I}}(\beta,p)}{2+\gamma_{\mathbf{I}}(\beta,p)}},\quad \mathfrak{L}(\mathbf{I})=\prod_{\mathbf{i}\in\mathbf{I}}\mathcal{L}_{\mathbf{i}}^{\frac{1}{\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})}}.$$

Here we have also used that $1/\gamma_{\mathbf{I}}(\beta,p)=\sum_{\mathbf{i}\in\mathbf{I}}1/\boldsymbol{\beta}_{\mathbf{i}}(\mathbf{I})$. The assertion of the theorem follows now from Theorem 1. ∎

5. Appendix.

5.1. *Proof of Proposition 1.* 1^0. Note that $\mathbf{M}(z)=\mathbf{M}(|z|)$ since $\mathbf{M}$ is symmetric that implies

$$\chi_r(y)=n^{-1}\sum_{j=1}^{n}\Big[M_r\big(\vec{\rho}\,(Y_j,y)\,\big)-\mathbb{E}_g^{(n)}\Big\{M_r\big(\vec{\rho}(Y_j,y)\big)\Big\}\Big],$$

where $\vec{\rho}:\mathbb{R}^s\times\mathbb{R}^s\to\mathbb{R}^s$ is given by $\vec{\rho}\big(z,z'\big)=\big(|z_1-z'_1|,\dots,|z_s-z'_s|\big)$.

We conclude that considered family of random fields obeys the structural assumption introduced in Section 4.4. of Lepski (2012), with $d=s$, $\mathbb{X}_1^d=\bar{\mathbb{X}}_1^d=\mathbb{R}^s$ and $\rho_l:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is given by $|z-z'|$ for any $l=\overline{1,s}$. It implies in particular that $\mathbb{R}^s$ is equipped with the metric ϱ_s generated by the supremum norm, i.e. $\varrho_s=\max_{l=\overline{1,s}}\rho_l$. We remark also that in our case $K(u)=\prod_{l=1}^{s}\mathbf{M}(u_l),\ u\in\mathbb{R}^s,\ g\equiv1$ and $\gamma_l=1, l=\overline{1,s}$.

To get the assertion of Proposition 1 we will apply Theorem 9 in Lepski (2012) on $\mathcal{R}_n(s):=[1/n,1]^s$. Note that obviously $\widetilde{\mathcal{R}}_n\subseteq\mathcal{R}_n(s)$. Thus, we have to check the assumptions of the latter theorem and to match the notations used in the present paper and in Lepski (2012).

First we note that since $\mathbf{M}$ satisfies Assumption 1 Assumption 9 (**i**) is obviously fulfilled with $L_1=(3s/2)(\mathrm{m}_\infty)^{s-1}L$. Moreover Assumption 9 (**ii**) holds because $g\equiv1$.

Thus, Assumption 9 is checked.

Consider the collection of closed cubs $\mathbb{B}_{\frac12}(\mathbf{j})=\{z\in\mathbb{R}^s:\ \ \varrho_s(z,\mathbf{j})\le1\}$, $\mathbf{j}\in\mathbb{Z}^s$, and let $\mathfrak{E}_{\mathbf{j}}(\delta),\ \delta>0$ denote the metric entropy of $\mathbb{B}_{\frac12}(\mathbf{j})$ measured in the metric ϱ_s.

Obviously $\Big\{\mathbb{B}_{\frac12}(\mathbf{j}),\ \mathbf{j}\in\mathbb{Z}^d\Big\}$ is a countable cover of $\mathbb{R}^s$ and each member of this collection is totally bounded (even compact) subset of $\mathbb{R}^s$. It is easily seen that

$$\mathrm{card}\left(\Big\{\mathbf{k}\in\mathbb{Z}^s:\ \mathbb{B}_{\frac12}(\mathbf{j})\cap\mathbb{B}_{\frac12}(\mathbf{k})\neq\emptyset\Big\}\right)\le3^s,\ \ \forall\mathbf{j}\in\mathbb{Z}^s.$$

Using the terminology of Lepski (2012) we can say that $\Big\{\mathbb{B}_{\frac12}(\mathbf{j}),\ \mathbf{j}\in\mathbb{Z}^d\Big\}$ is 3^s-totally bounded cover of $\mathbb{R}^s$. Moreover, $\mathfrak{E}_{\mathbf{j}}(\delta)=s\big[\ln(1/\delta)\big]_+$ for any $\delta>0$ and any $\mathbf{j}\in\mathbb{Z}^s$. All saying above allows us to assert that Assumption 7 (**i**) is fulfilled with $\mathbf{I}=\mathbb{Z}^s$, $\mathrm{X}_{\mathbf{j}}=\mathbb{B}_{\frac12}(\mathbf{j})$, $N=1.5s$ and $R=1$. It remains to note that Assumption 7 (**ii**) is automatically fulfilled in our case since $g\equiv1$.

Also we note that for any $\mathbf{j},\mathbf{k}\in\mathbb{Z}^s$ satisfying $\mathbb{B}_{\frac12}(\mathbf{j})\cap\mathbb{B}_{\frac12}(\mathbf{k})=\emptyset$ one has

$$\inf_{x\in\mathbb{B}_{\frac12}(\mathbf{j})}\ \inf_{y\in\mathbb{B}_{\frac12}(\mathbf{k})}\varrho_s(x,y)\ge1$$

and, therefore, Assumption 11 is checked with $\mathfrak{t}=1$. At last we have for any $n\geq 1$

$$\sup_{r\in\mathcal{R}_n(s)}\sup_{u\notin(0,1]^s}\left|\prod_{l=1}^{s}\mathbf{M}(u_l/r_l)\right|=0,$$

since $\operatorname{supp}(\mathbf{M})\subseteq[-1/2,1/2]$. Hence, the condition (4.24) of Theorem 9 is fulfilled as well that completes the verification of the assumptions of the theorem.

2^0. Let us match the notations. First, in our case $\mathbf{n_1}=\mathbf{n_2}=n$. Since $Y_j,\ j\geq 1$, are identically distributed the quantity denoted $F_{\mathbf{n_2}}\big(r,\bar{x}^{(d)}\big)$ is given now by $G(r,y)=\int_{\mathbb{R}^s}|M_r(x-y)|\mathbf{g}(x)\mathrm{d}x$ and, therefore, is independent on n. Here we have taken into account that $\bar{x}^{(d)}\in\mathbb{X}^d=\mathbb{R}^s$.

It is easily seen that

$$G_n:=\sup_{r\in[1/n,1]^s}\|G(r,\cdot)\|_\infty\leq\min\Big[\mathrm{m}_1^s\|\mathbf{g}\|_\infty,\mathrm{m}_\infty^s n^s\Big]. \tag{5.1}$$

It yields, in particular, that $F_{\mathbf{n_2}}=G_n\leq\mathrm{m}_1^s\|\mathbf{g}\|_\infty$ for any $n\geq 1$.

Choosing in Theorem 9 $q=p,\ v=2p+2,\ z=1$ and remembering that $\bar{x}^{(d)}=y$, we have

$$\widehat{\mathcal{U}}^{(v,z,p)}\big(n,r,\bar{x}^{(d)}\big)\leq\gamma_p\big(s,\mathrm{m}_\infty\big)\sqrt{\frac{\bar{G}(r)\ln(n)}{nV_r}},$$

for any $\bar{x}^{(d)}=y\in\mathbb{R}^s$ and any $r\in\widetilde{\mathcal{R}}_n(s)\subseteq\mathcal{R}_n(s)$. To get this assertion we have used that $G_n\leq(\mathrm{m}_\infty n)^s$ in view of (5.1).

At last, taking into account that the right hand side of the latter inequality is independent on y, we deduce from Theorem 9 that for any $p\geq 1$

$$\mathbb{E}^{(n)}_{\mathbf{g}}\left\{\sup_{r\in\widetilde{\mathcal{R}}_n(s)}\left[\big\|\chi_r\big\|_\infty-\gamma_p\big(s,\mathrm{m}_\infty\big)\sqrt{\frac{\bar{G}(r)\ln(n)}{nV_r}}\right\}^p_+\leq c_1(p,s)\big[1\vee\mathrm{m}_1^s\|\mathbf{g}\|_\infty\big]^{\frac{p}{2}}n^{-\frac{p}{2}}+c_2(p,s)n^{-p},$$

where $c_1(p,s)=2^{7p/2+5}3^{p+5s+4}\Gamma(p+1)\pi^p\big(s,\mathrm{m}_\infty\big)$ and $c_2(p,s)=2^{p+1}3^{5s}$. Here we have also used that $G_n\leq\mathrm{m}_1^s\|\mathbf{g}\|_\infty$ in view of (5.1) that implies $\widehat{F}_{\mathbf{n_2}}\leq 1\vee\mathrm{m}_1^s\|\mathbf{g}\|_\infty$. ∎

5.2. *Proof of Proposition 2.* First, noting that $\gamma_p\big(s,\mathrm{m}_\infty\big)\sqrt{\mathfrak{a}}=1/2$ we obtain from Proposition 1 that

$$\mathbb{E}^{(n)}_{\mathbf{g}}\left\{\sup_{r\in\widetilde{\mathcal{R}}^{(a)}_n(s)}\left(\big\|\chi_r\big\|_\infty-\frac{1}{2}\sqrt{\bar{G}(r)}\right)\right\}^p_+\leq c_n, \tag{5.2}$$

where we have put for brevity $c_n=c_1(p,s)\big[1\vee\mathrm{m}_1^s\|\mathbf{g}\|_\infty\big]^{\frac{p}{2}}n^{-\frac{p}{2}}+c_2(p,s)n^{-p}$. Next, putting $\bar{\chi}_r(y)=\Upsilon_r(y)-\mathbb{E}^n_{\mathbf{g}}\Upsilon_r(y)$ we have in view if (5.2)

$$\mathbb{E}^{(n)}_{\mathbf{g}}\left\{\sup_{r\in\widetilde{\mathcal{R}}^{(a)}_n(s)}\left(\big\|\bar{\chi}_r\big\|_\infty-\frac{1}{2}\sqrt{\bar{G}(r)}\right)\right\}^p_+\leq c_n. \tag{5.3}$$

To get the latter result we remarked that if $\mathbf{M}$ satisfies Assumption 1 then $|\mathbf{M}|$ satisfies it as well and, therefore, Proposition 1 is applicable to the process $\bar{\chi}_r(\cdot)$. It remains to note that the function $\bar{G}(\cdot)$ is the same for both processes $\chi_r(\cdot)$ and $\bar{\chi}_r(\cdot)$. We also note that

$$G(r)=\sup_{y\in\mathbb{R}^s}\Big\{\mathbb{E}^{(n)}_{\mathbf{g}}\Upsilon_r(y)\Big\}$$

and, therefore, for any $r\in(0,1]^s$ one has

$$\bar{G}(r)=1\vee\left\|\mathbb{E}^{(n)}_{\mathbf{g}}\Upsilon_r\right\|_\infty\leq 1\vee\|\Upsilon_r\|_\infty+\|\bar{\chi}_r\|_\infty\,, \tag{5.4}$$

where we have used the obvious inequality $\big|||x||\vee||z||-||y||\vee||z||\big|\leq||x-y||$ being true for any normed vector space.

Hence, putting $\zeta_n(\mathfrak{a})=\sup_{r\in\mathcal{R}^{(\mathfrak{a})}_n(s)}\left[\|\bar{\chi}_r\|_\infty-\frac{1}{2}\sqrt{\bar{G}(r)}\right]_+$ we obtain for any $r\in\mathcal{R}^{(\mathfrak{a})}_n(s)$

$$\bar{G}(r)\leq\frac{1}{2}\sqrt{\bar{G}(r)}+1\vee\|\Upsilon_r\|_\infty+\zeta_n(\mathfrak{a}).$$

It yields $\left[\bar{G}(r)-2\left(1\vee\|\Upsilon_r\|_\infty\right)\right]_+\leq 2\zeta_n(\mathfrak{a})$ and we have in view of (5.3)

$$\mathbb{E}^{(n)}_{\mathbf{g}}\left\{\sup_{r\in\mathcal{R}^{(\mathfrak{a})}_n(s)}\left[\bar{G}(r)-2\left(1\vee\|\Upsilon_r\|_\infty\right)\right]\right\}_+^p\leq 2^pc_n.$$

Similarly to (5.4) we have

$$1\vee\|\Upsilon_r\|_\infty\leq\bar{G}(r)+\|\bar{\chi}_r\|_\infty\leq(3/2)\bar{G}(r)+\zeta_n(\mathfrak{a})$$

and, therefore $\left[1\vee\|\Upsilon_r\|_\infty-(3/2)\bar{G}(r)\right]_+\leq\zeta_n(\mathfrak{a})$. Thus, we get from (5.3)

$$\mathbb{E}^{(n)}_{\mathbf{g}}\left\{\sup_{r\in\mathcal{R}^{(\mathfrak{a})}_n(s)}\left[1\vee\|\Upsilon_r\|_\infty-(3/2)\bar{G}(r)\right]\right\}_+^p\leq c_n.$$

∎

5.3. *Proof of Lemma 1.* We have in view of Fubini theorem for any $x_{\mathbf{I}}\in\mathbb{R}^{\mathbf{I}}$

$$\begin{aligned}
s^*_{h_{\mathbf{I}},\eta_{\mathbf{I}}}(x_{\mathbf{I}}) &= \int_{\mathbb{R}^{|\mathbf{I}|}}\left[K_{h_{\mathbf{I}}}\star K_{\eta_{\mathbf{I}}}\right](t_{\mathbf{I}}-x_{\mathbf{I}})f_{\mathbf{I}}(t_{\mathbf{I}})\mathrm{d}t_{\mathbf{I}}=\int_{\mathbb{R}^{|\mathbf{I}|}}\left[\int_{\mathbb{R}^{|\mathbf{I}|}}K_{\eta_{\mathbf{I}}}(y_{\mathbf{I}})K_{h_{\mathbf{I}}}(t_{\mathbf{I}}-x_{\mathbf{I}}-y_{\mathbf{I}})\mathrm{d}y_{\mathbf{I}}\right]f_{\mathbf{I}}(t_{\mathbf{I}})\mathrm{d}t_{\mathbf{I}}\\
&= \int_{\mathbb{R}^{|\mathbf{I}|}}K_{\eta_{\mathbf{I}}}(z_{\mathbf{I}}-x_{\mathbf{I}})\left[\int_{\mathbb{R}^{|\mathbf{I}|}}K_{h_{\mathbf{I}}}(t_{\mathbf{I}}-z_{\mathbf{I}})f_{\mathbf{I}}(t_{\mathbf{I}})\mathrm{d}t_{\mathbf{I}}\right]\mathrm{d}y_{\mathbf{I}}\\
&= s_{h_{\mathbf{I}}}(x_{\mathbf{I}})+\int_{\mathbb{R}^{|\mathbf{I}|}}K_{\eta_{\mathbf{I}}}(z_{\mathbf{I}}-x_{\mathbf{I}})\left[\int_{\mathbb{R}^{|\mathbf{I}|}}K_{h_{\mathbf{I}}}(t_{\mathbf{I}}-z_{\mathbf{I}})\left\{f_{\mathbf{I}}(t_{\mathbf{I}})-f_{\mathbf{I}}(z_{\mathbf{I}})\right\}\mathrm{d}t_{\mathbf{I}}\right]\mathrm{d}z_{\mathbf{I}}.
\end{aligned}$$

Therefore,

$$\left\|s^*_{h_{\mathbf{I}},\eta_{\mathbf{I}}}-s_{\eta_{\mathbf{I}}}\right\|_{\mathbf{I},\infty}\leq b_{h_{\mathbf{I}}}\int_{\mathbb{R}^{|\mathbf{I}|}}\left|K_{\eta_{\mathbf{I}}}(y_{\mathbf{I}})\right|\mathrm{d}y_{\mathbf{I}}\leq \mathrm{k}_1^d b_{h_{\mathbf{I}}}.$$

∎

5.4. *Proof of Lemma 2.* The proof of the lemma is completely based on application of Propositions 1–2 and Corollary 1.

Proof of (**i**). Remind that $\zeta(h,\mathcal{P})=\sup_{\mathbf{I}\in\mathcal{P}}\|\xi_{h_{\mathbf{I}}}\|_{\mathbf{I},\infty}$ and

$$\zeta_n=\sup_{\eta\in\mathcal{H}_n}\sup_{\mathcal{P}\in\mathfrak{P}}\Big[\zeta\big(\eta,\mathcal{P}\big)-\Lambda A_n(\eta,\mathcal{P})\Big]_+.$$

Then, we have

$$\Big[\mathbb{E}_f^{(n)}\big(\zeta_n\big)^{2q}\Big]^{\frac{1}{2q}}=\sum_{\mathcal{P}\in\mathfrak{P}}\sum_{\mathbf{I}\in\mathcal{P}}\left(\mathbb{E}_f^{(n)}\bigg\{\sup_{\eta_{\mathbf{I}}\in\mathcal{H}_n^{(\mathfrak{a}_{\mathbf{I}})}(|\mathbf{I}|)}\bigg[\|\xi_{\eta_{\mathbf{I}}}\|_{\mathbf{I},\infty}-\gamma_{2q}\big(|\mathbf{I}|,\mathrm{k}_\infty\big)\sqrt{\frac{\bar{s}_n\ln(n)}{nV_{\eta_{\mathbf{I}}}}}\bigg]\bigg\}_+^{2q}\right)^{\frac{1}{2q}},\tag{5.5}$$

where we have put $\mathcal{H}_n^{(\mathfrak{a}_{\mathbf{I}})}(|\mathbf{I}|)=\big\{\eta_{\mathbf{I}}\in(0,1]^{|\mathbf{I}|}:\ nV_{\eta_{\mathbf{I}}}\geq\mathfrak{a}_{\mathbf{I}}^{-1}\ln(n)\big\}$ and $\mathfrak{a}_{\mathbf{I}}=\big[2\gamma_{2q}\big(\mathbf{I},\mathrm{k}_\infty\big)\big]^{-2}$. To get the latter result we have used first that $A_n(\eta,\mathcal{P})=\sup_{\mathbf{I}\in\mathcal{P}}\sqrt{\frac{\bar{s}_n\ln(n)}{nV_{\eta_{\mathbf{I}}}}}$ and the trivial inequality $\big[\sup_i x_i-\sup_i y_i\big]_+\leq\sup_i[x_i-y_i]_+$. Next we have used that $\Lambda=\sup_{\mathcal{P}\in\mathfrak{P}}\sup_{\mathbf{I}\in\mathcal{P}}\gamma_{2q}\big(|\mathbf{I}|,\mathrm{k}_\infty\big)$. At last we have used that $\eta\in\mathcal{H}_n$ implies $\eta_{\mathbf{I}}\in\mathcal{H}_n^{(\mathfrak{a}_{\mathbf{I}})}(|\mathbf{I}|)$ for any $\mathbf{I}\in\mathcal{I}_d$ in view of the definition of $\mathfrak{a}^*$.

Note that for any for any $\mathbf{I}\in\mathcal{I}_d$ and any $\eta_{\mathbf{I}}\in(0,1]^{|\mathbf{I}|}$

$$\bar{s}\geq1\vee\bigg\|\int_{\mathbb{R}^{\mathbf{I}}}\big|K_{\eta_{\mathbf{I}}}\big(t_{\mathbf{I}}-\cdot\big)\big|f_{\mathbf{I}}\big(t_{\mathbf{I}}\big)\mathrm{d}t_{\mathbf{I}}\bigg\|_{\mathbf{I},\infty}=:\bar{F}_{\mathbf{I}}(\eta).$$

We conclude that Proposition 1 is applicable with $\chi_r=\xi_{\eta_{\mathbf{I}}}$, $\mathbf{M}=\mathbf{K}$, $p=2q$, $s=|\mathbf{I}|$, $\mathfrak{a}=\mathfrak{a}_{\mathbf{I}}$, $\bar{G}=\bar{F}_{\mathbf{I}}$ and the assertion (**i**) follows with

$$\mathbf{c_1}\big(2q,d,\mathbf{K},\mathbf{f}\big)=\sum_{\mathcal{P}\in\mathfrak{P}}\sum_{\mathbf{I}\in\mathcal{P}}\Big[c_1\big(2q,|\mathbf{I}|\big)\big[1\vee\mathrm{k}_1^{|\mathbf{I}|}\mathbf{f}\big]^q+c_2\big(2q,|\mathbf{I}|\big)\Big].$$

Proof of (**ii**). Put for any $h\in\mathcal{H}_n$ and $\mathbf{I}\in\mathcal{I}_d$

$$s_{\mathbf{I}}\big(h_{\mathbf{I}}\big)=\bigg\|\int_{\mathbb{R}^{\mathbf{I}}}\big|K_{h_{\mathbf{I}}}\big(t_{\mathbf{I}}-\cdot\big)\big|f_{\mathbf{I}}\big(t_{\mathbf{I}}\big)\mathrm{d}t_{\mathbf{I}}\bigg\|_{\mathbf{I},\infty}\qquad f_{\mathbf{I},n}\big(h_{\mathbf{I}}\big)=\bigg\|n^{-1}\sum_{i=1}^{n}\big|K_{h_{\mathbf{I}}}\left(X_{\mathbf{I},i}-\cdot\right)\big|\bigg\|_{\mathbf{I},\infty}.$$

We have similarly to (5.5) $[\bar{s}_n-\bar{\mathbf{f}}_n]_+\leq\sup_{\mathbf{I}\in\mathcal{I}_d}\sup_{h_{\mathbf{I}}\in\mathcal{H}_n^{(\mathfrak{a}_{\mathbf{I}})}(|\mathbf{I}|)}\big[s_{\mathbf{I}}\big(h_{\mathbf{I}}\big)-2\mathbf{f}_{\mathbf{I},n}\big(h_{\mathbf{I}}\big)\big]_+$ and hence

$$\Big[\mathbb{E}_f^{(n)}\left[\bar{s}_n-\bar{\mathbf{f}}_n\right]_+^{2q}\Big]^{\frac{1}{2q}}\leq\sum_{\mathbf{I}\in\mathcal{I}_d}\left(\mathbb{E}_f^{(n)}\bigg\{\sup_{h_{\mathbf{I}}\in\mathcal{H}_n^{(\mathfrak{a}_{\mathbf{I}})}(|\mathbf{I}|)}\Big[s_{\mathbf{I}}\big(h_{\mathbf{I}}\big)-2\mathbf{f}_{\mathbf{I},n}\big(h_{\mathbf{I}}\big)\Big]\bigg\}_+^{2q}\right)^{\frac{1}{2q}},$$

The assertion (**ii**) follows now from the second statement of Proposition 2 with

$$\mathbf{c_2}\big(2q,d,\mathbf{K},\mathbf{f}\big)=\sum_{\mathbf{I}\in\mathcal{I}_d}\Big[c_1'\big(2q,|\mathbf{I}|\big)\big[1\vee\mathrm{k}_1^{|\mathbf{I}|}\mathbf{f}\big]^q+c_2'\big(2q,|\mathbf{I}|\big)\Big].$$

Proof of (**iii**). We have $[\bar{\mathbf{f}}_n-3\bar{s}_n]_+\leq2\sup_{\mathbf{I}\in\mathcal{I}_d}\sup_{h_{\mathbf{I}}\in\mathcal{H}_n^{(\mathfrak{a}_{\mathbf{I}})}(|\mathbf{I}|)}\big[\mathbf{f}_{\mathbf{I},n}\big(h_{\mathbf{I}}\big)-(3/2)s_{\mathbf{I}}\big(h_{\mathbf{I}}\big)\big]_+$ and hence

$$\Big[\mathbb{E}_f^{(n)}\left[\bar{\mathbf{f}}_n-3\bar{s}_n\right]_+^{2q}\Big]^{\frac{1}{2q}}\leq2\sum_{\mathbf{I}\in\mathcal{I}_d}\left(\mathbb{E}_f^{(n)}\bigg\{\sup_{h_{\mathbf{I}}\in\mathcal{H}_n^{(\mathfrak{a}_{\mathbf{I}})}(|\mathbf{I}|)}\Big[\mathbf{f}_{\mathbf{I},n}\big(h_{\mathbf{I}}\big)-(3/2)s_{\mathbf{I}}\big(h_{\mathbf{I}}\big)\Big]\bigg\}_+^{2q}\right)^{\frac{1}{2q}},$$

The assertion (**iii**) follows now from the first assertion of Proposition 2 with

$$\mathbf{c_3}\big(2q,d,\mathbf{K},\mathbf{f}\big)=2\sum_{\mathbf{I}\in\mathcal{I}_d}\Big[c_1\big(2q,|\mathbf{I}|\big)\big[1\vee\mathrm{k}_1^{|\mathbf{I}|}\mathbf{f}\big]^q+c_2\big(2q,|\mathbf{I}|\big)\Big].$$

Proof of (**iv**). Note that

$$\begin{aligned}\bar{\mathfrak{f}}_n &:= 2d^2\mathrm{k}_1^d\big(\bar{\mathbf{f}}_n\big)^{\lfloor d^2/4\rfloor+1}\left[\max\left\{\bar{\mathbf{f}}_n,\mathrm{k}_1^2\mathbf{f}\right\}\right]^{d-1}\\ &\le \beta\left[\big(\mathbf{f}_n\big)^{\lfloor d^2/4\rfloor+d}+\big(1+\mathrm{k}_1^2\mathbf{f}\big)^{d-1}\big(\mathbf{f}_n\big)^{\lfloor d^2/4\rfloor+1}+\big(\mathbf{f}_n\big)^{d-1}+\big(1+\mathrm{k}_1^2\mathbf{f}\big)^{d-1}\right], \end{aligned}\tag{5.6}$$

where we have used $\mathrm{k}_1\ge 1$ and put $\beta=2d^2\mathrm{k}_1^d2^{\lfloor d^2/4\rfloor+d}$. Thus, to get the assertion (**iv**) it suffices to bound from above $\mathbb{E}_f\big(\mathbf{f}_n\big)^p,\ p\ge 1$. We obviously have

$$\mathbf{f}_n\le\sum_{\mathbf{I}\in\mathcal{I}_d}\sup_{h_{\mathbf{I}}\in\mathcal{H}_n^{(\mathfrak{a}_{\mathbf{I}})}(|\mathbf{I}|)}\left\|n^{-1}\sum_{i=1}^n\left|K_{h_{\mathbf{I}}}\left(X_{\mathbf{I},i}-\cdot\right)\right|\right\|_{\mathbf{I},\infty},$$

and using Corollary 1 we get for $p\ge 1$

$$\left[\mathbb{E}_f^{(n)}\big(\mathbf{f}_n\big)^p\right]^{\frac1p}\le\sum_{\mathbf{I}\in\mathcal{I}_d}\left[1\vee\mathrm{k}_1^{|\mathbf{I}|}\mathbf{f}\right]^{\frac12}\left[\gamma_p\big(|\mathbf{I}|,\mathrm{k}_\infty\big)+\left\{c_1(p,|\mathbf{I}|)+c_2(p,|\mathbf{I}|)\right\}^{\frac1p}\right].\tag{5.7}$$

The assertion (**iv**) follows now from (5.6) and (5.7). ∎

5.5. *Proof of Lemma 4.* The proof of the lemma is based on the embedding theorem for anisotropic Nikolskii classes which we formulate below.

Let $(\alpha,r)\in(0,\infty)^s\times[1,\infty]^s$ be such that $\varkappa=1-\sum_{l=1}^s(\alpha_lr_l)^{-1}>0$ and let $Q\in(0,\infty)^s$. Then there exists $\mathbf{c}>0$ completely determined by α,r and s such that

$$\mathbb{N}_{r,s}\big(\alpha,Q\big)\subseteq\mathbb{N}_{\infty,s}\big(\boldsymbol{\alpha},\mathbf{c}Q\big),\tag{5.8}$$

where $\boldsymbol{\alpha}=(\boldsymbol{\alpha}_1,\dots\boldsymbol{\alpha}_s\ )$, $\boldsymbol{\alpha}_j=\varkappa\alpha_j\varkappa_j^{-1}$ and $\varkappa_j=1-\sum_{l=1}^s\big(r_l^{-1}-r_j^{-1}\big)\alpha_l^{-1}$.

The inclusion (5.8) is a particular case of Theorem 6.9 in Nikol'skii (1977), with $p'=\infty$. We remark that $\mathbb{N}_{\infty,s}\big(\boldsymbol{\alpha},\mathcal{Q}\big)$ is anisotropic Hölder class of functions.

Let $\mathbf{E}_i,\ i=\overline{1,s}$ be the family of $s\times s$ matrices where $\mathbf{E}_i=(\mathbf{e}_1,\dots,\mathbf{e}_i,\mathbf{0}\dots,\mathbf{0})$ and let $\mathbf{E}_0$ is zero matrix. Putting $K(u)=\prod_{l=1}^s\mathbf{K}(u_l),\ u_l\in\mathbb{R}^s$, we get for any $\eta\in(0,\infty)^s$ and any $z\in\mathbb{R}^s$

$$|\mathcal{B}_{\eta,g}(z)|=\left|\int_{\mathbb{R}^s}K(u)\left[g(z+u\eta)-g(z)\right]\mathrm{d}u\right|\le\sum_{i=1}^s\left|\int_{\mathbb{R}^s}K(u)\Big[g\big(z+\eta\mathbf{E}_iu\big)-g\big(z+\eta\mathbf{E}_{i-1}u\big)\Big]\mathrm{d}u\right|.$$

We note that the all components of the vectors $z+\eta\mathbf{E}_iu$ and $z+\eta\mathbf{E}_{i-1}u$ except i-th coordinate coincide. Hence using Taylor expansion we obtain any $\eta\in(0,\infty)^s$ and $z\in\mathbb{R}^s$ in view of (5.8)

$$\left|\int_{\mathbb{R}^s}K(u)\Big[g\big(z+\eta\mathbf{E}_iu\big)-g\big(z+\eta\mathbf{E}_{i-1}u\big)\Big]\mathrm{d}u\right|\le\mathbf{c}Q_i\eta_i^{\boldsymbol{\alpha}_i}\int_{\mathbb{R}^s}|K(u)||u|^{\boldsymbol{\alpha}_i}\mathrm{d}u\le\mathrm{k}_1^s\mathbf{c}Q_i\eta_i^{\boldsymbol{\alpha}_i}.$$

To get the last inequality we have taken into account (3.1) and used that $\mathbf{K}$ is supported on $[-1/2,1/2]$. It is worth mentioning that $\mathbf{c}$ as a function of α is bounded on any bounded domain of $(0,\infty)^s$. Since the right hand side of the latter inequality is independent of z we come to the assertion of the lemma. ∎

References.

Akakpo, N. (2012). Adaptation to anisotropy and inhomogeneity via diadic piecewise polynomial selection. *Math. Methods Statist.* **21**, 1–28.

Birgé, L. (2008). Model selection for density estimation with $\mathbb{L}_2$–loss. arXiv:0808.1416v2, http://arxiv.org

Bertin, K. (2005). Sharp adaptive estimation in sup-norm for d-dimensional Holder classes. *Math. Methods Statist.* **14**, 267–298.

Bretagnolle, J. and Huber, C. (1979). Estimation des densités: risque minimax. *Z. Wahrsch. Verw. Gebiete* **47**, 119–137.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation. The* $\mathbb{L}_1$ *View.* Wiley & Sons, New York.

Devroye, L. and Lugosi, G. (1996). A universally acceptable smoothing factor for kernel density estimation. *Ann. Statist.* **24**, 2499–2512.

Devroye, L. and Lugosi, G. (1997). Nonasymptotic universal smoothing factors, kernel complexity and Yatracos classes. *Ann. Statist.* **25**, 2626–2637.

Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation.* Springer, New York.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508–539.

Efroimovich, S. Yu. (1986). Non-parametric estimation of the density with unknown smoothness. *Ann. Statist.* **36**, 1127–1155.

Efroimovich, S. Yu. (2008). Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *Theory Probab. Appl.* **30**, 557–568.

Giné E. and Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with application to adaptive estimation. *Probability Theory and Related Fields* **143**, 569–596.

Giné E. and Nickl, R. (2010). Adaptive estimation of the distribution function and its density in sup-norm loss by wavelets and spline projections. *Bernoulli* **16**, 1137–1163.

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Stat.* **39, 3**, 1608–1632.

Golubev, G.K. (1992). Non-parametric estimation of smooth probability densities. *Probl. Inform. Transm.* **1**, 52–62.

Hasminskii, R. and Ibragimov, I. (1990). On density estimation in the view of Kolmogorov's ideas in approximation theory. *Ann. Statist.* **18**, 999–1010.

Ibragimov, I. A. and Khasminski, R. Z. (1980). An estimate of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **98**, 61–85 (in Russian).

Ibragimov, I. A. and Khasminski, R. Z. (1981). More on estimation of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **108**, 72–88 (in Russian).

Jennrich, R. (1969). Asymptotic properties of non–linear least squares estimators. *Ann. Math. Statist.* **40**, 633–643.

Juditsky, A. and Lambert–Lacroix, S. (2004). On minimax density estimation on $\mathbb{R}$. *Bernoulli* **10**, 187-220.

Kerkyacharian, G., Lepski, O. and Picard, D. (2001). Nonlinear estimation in anisotropic multi–index denoising. *Probab. Theory Related Fields* **121**, 137–170.

Kerkyacharian, G., Lepski, O. and Picard, D. (2007). Nonlinear estimation in anisotropic multi–index denoising. Sparce case. *Probab. Theory Appl.* **52**, 150–171.

Kerkyacharian, G., Picard, D. and Tribouley, K. (1996). L^p adaptive density estimation. *Bernoulli* **2**, 229–247.

Lepski, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36**, 682–697.

Lepski, O.V. (1992). On problems of adaptive estimation in white Gaussian noise. *IN: Topics in nonparametric estimation,* **12**, Adv. Soviet Math, 187-220, Amer. math. Soc, Providence RI.

Lepski, O. (2012). Upper functions for positive random functionals. arXiv:1202.6615v1, http://arxiv.org.

Massart, P. (2007). *Concentration Inequalities and Model Selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. Lecture Notes in Mathematics, 1896. Springer, Berlin.

Mason, D. M. (2009). Risk bounds for kernel density estimators. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **363**, 66–104. Available at http://www.pdmi.ras.ru/znsl/

Nikol'skii, S. M. (1977). *Priblizhenie Funktsii Mnogikh Peremennykh i Teoremy Vlozheniya.* (in Russian). [Approximation of functions of several variables and imbedding theorems.] Nauka, Moscow, 1977.

Parzen, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Statist.* **33**, 1065-1076.

Reynaud-Bouret, P., Rivoirard, V. and Tuleau-Malot, C. (2011). Adaptive density estimation: a course of support? *J. Statist. Plann. Inference* **141**, 115–139.

Rigollet, Ph. (2006). Adaptive density estimation using the blockwise Stein method. *Bernoulli* **12**, 351–370.

Rigollet, Ph. and Tsybakov, A. B. (2007). Linear and convex aggregation of density estimators. *Math. Methods*

Statist. **16**, 260–280.

Rosenblatt, M (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832-837.

Samarov, A. and Tsybakov, A. (2007). Aggregation of density estimators and dimension reduction. *Advances in Statistical Modeling and Inference*, 233–251, Ser. Biostat., 3, World Sci. Publ., Hackensack, NJ.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, London.

Tsybakov, A. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann.Statist.* **26**, 2420–2469.

Laboratoire d'Analyse, Topologie, Probabilités
Université Aix-Marseille
39, rue F. Joliot-Curie
13453 Marseille, France
E-mail: lepski@cmi.univ-mrs.fr